Content Moderation

Getting Started

 Issue
 01

 Date
 2022-12-05





HUAWEI TECHNOLOGIES CO., LTD.

Copyright © Huawei Technologies Co., Ltd. 2022. All rights reserved.

No part of this document may be reproduced or transmitted in any form or by any means without prior written consent of Huawei Technologies Co., Ltd.

Trademarks and Permissions

NUAWEI and other Huawei trademarks are trademarks of Huawei Technologies Co., Ltd. All other trademarks and trade names mentioned in this document are the property of their respective holders.

Notice

The purchased products, services and features are stipulated by the contract made between Huawei and the customer. All or part of the products, services and features described in this document may not be within the purchase scope or the usage scope. Unless otherwise specified in the contract, all statements, information, and recommendations in this document are provided "AS IS" without warranties, guarantees or representations of any kind, either express or implied.

The information in this document is subject to change without notice. Every effort has been made in the preparation of this document to ensure accuracy of the contents, but all statements, information, and recommendations in this document do not constitute a warranty of any kind, express or implied.

Huawei Technologies Co., Ltd.

- Address: Huawei Industrial Base Bantian, Longgang Shenzhen 518129 People's Republic of China Website: https://www.huawei.com
- Email: <u>support@huawei.com</u>

Contents

1 Content Moderation Overview	1
2 Subscribing to the Service	2
3 Using Postman to Call Content Moderation	4
3.1 Moderation (Text)	4
3.2 Moderation (Image)	10
4 Using an SDK to Call Content Moderation	17
4.1 Moderation (Text)	17
4.2 Moderation (Image)	23

Content Moderation Overview

Content Moderation adopts image, text, and audio and video detection technologies that detect pornography and images and text violating related laws or regulations. This reduces potential business risks.

Content Moderation provides services through open application programming interfaces (APIs). You can obtain the inference result by calling APIs. It helps you build an intelligent service system and improves service efficiency. Currently, Content Moderation includes Moderation (Image) and Moderation (Text).

The following methods are available to call Content Moderation APIs:

• A software tool such as curl and Postman

These are good options if you are more comfortable writing code, HTTP requests, and API calls. For details, see **Using Postman to Call Content Moderation**.

• Software development kit (SDK)

SDKs for Java, Python, .NET, and GO are available for quick integration. For details, see **Using an SDK to Call Content Moderation**.

2 Subscribing to the Service

NOTE

This service is available only to enterprise users for now.

To subscribe to Content Moderation, perform the following steps:

- 1. Register a HUAWEI CLOUD account and complete real-name authentication.
- 2. Log in to the Content Moderation console and select a region. For details about the region where the Content Moderation service is available, see **Endpoints**.
- 3. In the lower right corner of the page, click **Customer Service** to subscribe to the Content Moderation (Text) service.
- 4. After a commercial service is successfully subscribed, the service is displayed in **My Services** on the **Service Management** page. In this case, you can call the API to use Content Moderation.
- 5. In the upper right corner, click **Prepay to Get Discounts**. On the displayed page, select your desired type and specifications, set other parameters, and click **Next**. Then confirm the information, click **Submit**, and pay the order to enable the service.

Buy Content Moderation Package ③ * Region CN North-Beijing4 The purchased resource package applies to only the selected region. It cannot be shared across regio Image Moderation-Antiporn Image Moderation-Political Figure Image Moderation-Politics and Terrorism Text Moderation Type Distortion Correction Image Clarity Detect Image Moderation-Ad Audio Moderation-Short 600 thousand API calls 1500 thousand API calls 3 million API calls 6 million API calls 15 million API calls Specification 30 million API calls 60 million API calls 120 million API calls 1 If resources in a package have been used up, subsequent usage of the resources will be billed according to standard pricing. 2 months 3 months 4 months 5 months 6 months 7 months 8 months 9 months 1 year 2 years Usage Duration 3 years - 1 + Purchase Quantity

Figure 2-1 Buy Content Moderation Package

Content Moderation supports both the pay-per-use and yearly/monthly billing modes. For details, see **Billing**. If you want to use the pay-per-use billing mode, see **Content Moderation Pricing Details**.

3 Using Postman to Call Content Moderation

3.1 Moderation (Text)

This section provides an example of how to use Postman to call Moderation (Text), helping you quickly get familiar with the service. The procedure is as follows:

Step 1: Subscribe to a Service: Subscribe to Moderation (Text) on the Content Moderation console.

Step 2: Configure a Custom Dictionary. Configure allowlist and blocklist dictionaries.

Step 3: Configure the Environment. Import the prepared configuration file into the development environment.

Step 4: Use a Token for Authentication. When you call an API, a token is required for authentication.

Step 5: Call a Service API. Call the service API. You can view the status code and error code at any time during the usage.

(Optional) Step 6: View the Number of API Calls. View call statistics and call details.

Step 1: Subscribe to a Service

Subscribe to Moderation (Text) by following the instructions in **Subscribing to the Service**.

Step 2: Configure a Custom Dictionary

You can configure allowlist and blocklist dictionaries for text moderation.

 If an allowlist dictionary is customized, the words in the allowlist dictionary will not be detected when you use text moderation.
 The procedure is as follows:

- a. Log in to the HUAWEI CLOUD and click **Console** in the upper right corner. The **Console** page is displayed.
- b. Click **Service List** and choose **AI** > **Content Moderation**. The **Content Moderation** management console is displayed.
- c. In the left navigation pane, choose Service List > Text Moderation. Click Configuration next to Custom Allowlist Dictionary. The Create Dictionary dialog box is displayed.

Figure 3-1 Create Dictionary

Create Dictio	nary	×
* Dictionary Name	Enter a dictionary nam	
* Add Keyword	You have entered 0 keywords.	
Enter a keyword.		
	OK Cancel	

- d. Enter a dictionary name and the keywords in the allowlist dictionary. Click **OK**.
 - Dictionary Name: A dictionary name contains a maximum of 40 characters. Only letters, digits, hyphens (-), and underscores (_) are allowed.
 - Add Keyword: A dictionary contains a maximum of 5000 keywords. Each keyword consists of a maximum of 40 characters and ends with a carriage return.
- If a blocklist dictionary is customized, the words in the blocklist dictionary will be detected when you use Text Moderation.

The procedure is as follows:

- a. Log in to the HUAWEI CLOUD and click **Console** in the upper right corner. The **Console** page is displayed.
- b. Click Service List and choose Al > Content Moderation. The Content Moderation management console is displayed.
- c. In the left navigation pane, choose Service List > Text Moderation. Click Configuration next to Custom Blocklist Dictionary. The Custom Dictionary page is displayed.
- d. Click **Create Dictionary** in the upper left corner. The **Create Dictionary** dialog box is displayed. Enter a dictionary name and the keywords in the blocklist dictionary.

- i. **Dictionary Name**: A dictionary name contains a maximum of 40 characters. Only letters, digits, hyphens (-), and underscores (_) are allowed.
- ii. **Add Keyword**: A dictionary contains a maximum of 5000 keywords. Each keyword consists of a maximum of 40 characters and ends with a carriage return.
- e. Click OK.

• The dictionary name cannot be modified.

Step 3: Configure the Environment

1. Download and install Postman.

Postman 7.24.0 is recommended.

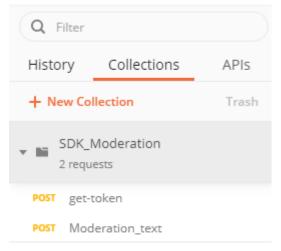
- 2. Download the Postman.
- 3. Import the configuration file.
 - a. Open and log in to Postman.
 - b. Choose **File** > **Import** in the upper left corner. In the displayed dialog, click **Upload Files** to import the configuration file.

Figure 3-2 Importing the configuration file

File Edit View Help			
H New T Import R	lunner		👪 My Workspace 👻 🌲 Invite
Q Filter		[DELETED] POST get-token	[DELETED] POST Image_tagging + ••
History Collections	APIs		
+ New Collection		IMPORT	
SDK_Moderation 2 requests		File Folder Link	Raw Text
		Drag	and drop Postman data or any of the formats
		OF	penAPI RAML GraphQL cURL WA
			OR
			Upload Files

4. After the configuration file is imported, it is displayed in the **Collections** directory in the left navigation pane.

Figure 3-3 Configuration Files



Step 4: Use a Token for Authentication

A token is a user's access credential, which includes user identities and permissions. When you call an API to access a cloud service, a token is required for identity authentication.

- 1. In the **Collections** directory in the navigation pane of Postman, click the **get-token** configuration file.
- 2. Click **Body**. Enter the username, password, and domain name.

NOTE

To obtain the username, password, and domain name, perform the following steps:

- 1. Log in to the management console.
- 2. Hover the cursor on the username and choose **My Credentials** from the dropdown list. Copy the IAM username and account name to **username** and **domainname**, and enter the IAM user password.

Q Filter History Collections APIs	Post get-token × + ····
+ New Collection Trash	
SDK_Moderation 2 requests	POST
POST get-token	none form-data x-www-form-urlencoded raw binary GraphQL JSON 🔻
P05T Moderation_text	<pre>1 * { 2 * "auth": { 3 * "identity": { 4 * "methods": [5 "password" 6], 7 * "password": { 8 * "username", 9 "name": "username", 10 "name": "xxxxxxxx", 11 * "domain": { 12 "name": "domainname" 13 } 14 } 15 } 16 }, 17 * "scope": { 18 * "project": { 19 "name": "cn-north-4" 20 } 21 } 22 } 23 }</pre>

Figure 3-4 Token-based authentication

Click Send to send a request, and obtain and copy the token.
 In Headers, obtain the value of X-Subject-Token, which is the token. The token is valid for 24 hours.

Figure 3-5 Obtaining a token

Q Filter		POST get-token	No Environment
History Collections	APIs	POST v https://iam.cn-north-4.myhuaweicloud.co	m/v3/auth/tokens Send 🔻
+ New Collection	Trash		
SDK Moderation		KEY	VALUE
2 requests		via ①	proxy A
POST get-token		Date (i)	Thu, 13 May 2021 09:10:21 GMT
POST Moderation_text		Pragma (i)	no-cache
		Server (3)	api-gateway
		Expires (3)	Thu, 01 Jan 1970 00:00:00 GMT
		Connection ③	Keep-Alive
		Content-Type ③	application/json; charset=UTF-8
		X-Request-Id (1)	e21f61aa7318d38dfd4dcc3ece08ed15
		Cache-Control ③	no-cache, no-store, must-revalidate
		Content-Length (3)	17825
		X-IAM-Trace-Id (3)	token_cn-north-4_null_2d2d2db31e68b0995c49a311ee21c340
		X-Frame-Options (i)	SAMEORIGIN
		X-Subject-Token ③	MIIVMQYJKoZlhvcNAQcCollVljCCFR4CAQExDTALBglghkgBZQMEAg
		X-XSS-Protection ③	1; mode=block;
		X-Download-Options ③	noopen
		X-Content-Type-Options ③	nosniff
		Strict-Transport-Security ③	max-age=31536000; includeSubdomains;

Step 5: Call a Service API

- 1. In the **Collections** directory in the navigation pane of Postman, click the **Moderation_image** configuration file.
- 2. Click **Headers** and copy the token value to **X-Auth-Token**.

Figure 3-6 Entering the token

Q Filter	POST get-token POST Moderation_text + •••						
History Collections APIs	Moderation_text						
+ New Collection Trash	POST v https://moderation.cn-north-4.myhuaweicloud.com/v1.0/moderation/text						
SDK_Moderation	POST						
2 requests	Params Authorization Headers (10) Body Pre-request Script Tests Settings						
POST get-token	Headers 🐵 8 hidden						
POST Moderation_text	KEY VALUE						
	Content-Type application/json						
	X-Auth-Token						
	Key Value						

- 3. Click **Body** and enter the text to be checked for the **text** parameter.
 - For details about other parameters in the body, see the Text Moderation API.

Figure 3-7 Modifying the parameters

Params	Authorization	Headers (10)	Body 鱼	Pre-re	equest Script	Tests	Settings	
none	form-data	• x-www-form-u	rlencoded	🖲 raw	binary	GraphQL	JSON	Ŧ
1 - {								
2 • 3 4 5 6 7 8 9 10 • 11 • 12 13 14 15 16 }		,	ed",					

4. Click **Send** to send the request and obtain the API call results.

ι	
	"result": {
	"detail": {},
	"suggestion": "pass"
	}
}	

(Optional) Step 6: View the Number of API Calls

Step 1 View the number of API calls. Choose **Service List > Text Moderation**, and view call statistics and call details. See **Figure 3-8**.

Figure 3-8 Call statistics

Call Statistics					
Daily Text Moderation API calls		Daily blocked text		Daily reviewed text	
1		1		1	
0.8		0.8		0.8	
0.6		0.6		0.6	
0.4		0.4		0.4	
0.2		0.2		0.2	
0 06-12 06-13 06-14 06-1 2020 2020 2020 202 Total cal Call Details		0 06-12 06-13 06-14 06-15 2020 2020 2020 2020 Total blocks	06-16 06-17 2020 2020	0 05-12 05-13 06-14 06-15 2020 2020 2020 2020 • Total reviews	05-16 05-17 2020 2020
Call Details				2020-06-12 - 20	020-06-17 × 首 Confirm Export
Date ‡	Call Type	Total ‡	Blocks \$	Reviews \$	Passes \$
2020-06-17	Text	0	0	0	0
2020-06-16	Text	0	0	0	0
2020-06-15	Text	0	0	0	0
2020-06-14	Text	0	0	0	0
2020-06-13	Text	0	0	0	0
2020-06-12	Text	0	0	0	0

- The **Call Statistics** section provides three statistical charts, displaying daily calls, daily blocked text, and daily reviewed text. You can set a time range to view the change of the number of API calls within the time range.
- The **Call Details** section lists the total number of calls, blocks, reviews, and passes every day. This helps you better understand service calls and their statuses.

NOTE

- block indicates that sensitive information is contained and the information is blocked.
- pass indicates that sensitive information is not contained and the information is approved.
- **review** indicates that manual review is required.
- **Step 2** View monitoring metrics. You can click **View Metric** to view historical data such as successful and failed calls on the Cloud Eye console. See **Figure 3-9**.

Figure 3-9 Viewing metrics

1h	3h 12h	24 hours	7d 📋	Auto Refresh							
Period Raw	data 👻										
Successfu	l Calls of Service ?				+ ~	Failed Calls of	f Service ⑦				$+$ \sim
Times/min					Max Min	Times/min					Max Min
1.2					0 0	1.2					0 0
0.9						0.9					
0.6						0.6					
0.3						0.3					
0						0					
13:50	14:02	14:14	14:26	14:38	14:50	13:50	14:02	14:14	14:26	14:38	14:50

3.2 Moderation (Image)

This section provides an example of how to use Postman to call Moderation (Image), helping you quickly get familiar with the service. The procedure is as follows:

Step 1: Subscribe to a Service. In the left navigation pane, choose **Service List** or **Service Management**, and subscribe to Moderation (Image).

(Optional) Step 2: Enable Object Storage Service (OBS) Authorization. If you want to use the data stored on OBS, enable OBS authorization.

(Optional) Step 3: Configure a Moderation Rule. Detect pornographic content in different scenarios.

Step 4: Configure the Environment. Import the prepared configuration file into the development environment.

Step 5: Use a Token for Authentication. When you call an API, a token is required for authentication.

Step 6: Call a Service API. Call the service API. You can view the status code and error code at any time during the usage.

(Optional) Step 7: View the Number of API Calls. View call statistics and call details.

Step 1: Subscribe to a Service

Subscribe to Moderation (Image) by following the instructions in **Subscribing to the Service**.

(Optional) Step 2: Enable Object Storage Service (OBS) Authorization

If you want to use the data stored on OBS, enable OBS authorization. The procedure is as follows:

- 1. Log in to the management console and choose **Service List** > **Content Moderation**. In the navigation pane, choose **Service Management**.
- 2. Enable **OBS Authorization**.

NOTE

- The region of OBS must be consistent with that of Content Moderation.
- OBS authorization must be performed by an account, not by an IAM user.
- Once enabled, OBS authorization cannot be disabled.

(Optional) Step 3: Configure a Moderation Rule

You can configure review rules to detect pornographic content in different scenarios. For details, see **Configuring Moderation Rules**.

Step 4: Configure the Environment

1. Download and install Postman.

Postman 7.24.0 is recommended.

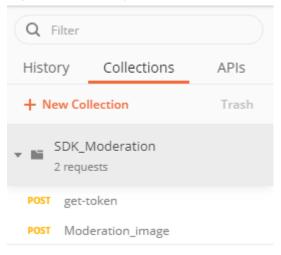
- 2. Download the Postman.
- 3. Import the configuration file.
 - a. Open and log in to Postman.
 - b. Choose **File** > **Import** in the upper left corner. In the displayed dialog, click **Upload Files** to import the configuration file.

File Edit View Help									
New T Import						🔡 My	Workspace	▼	Invite
Q Filter		[DELETED] P	DST get-toke	n	• [D	ELETED] POST	Image_tagging	; •	+ ••
History Collections	APIs	IMPORT		_		_	_		
+ New Collection		IMPORT							
SDK_Moderation		File	Folder	Link	Raw Te	ĸt			
					end droj	RAML	data or any GraphQL OR Upload Files	r of the fo	ormats WA

Figure 3-10 Importing the configuration file

4. After the configuration file is imported, it is displayed in the **Collections** directory in the left navigation pane.

Figure 3-11 Configuration Files



Step 5: Use a Token for Authentication

A token is a user's access credential, which includes user identities and permissions. When you call an API to access a cloud service, a token is required for identity authentication.

1. In the **Collections** directory in the navigation pane of Postman, click the **get-token** configuration file.

2. Click **Body**. Enter the username, password, and domain name.

NOTE

To obtain the username, password, and domain name, perform the following steps:

- 1. Log in to the management console.
- 2. Hover the cursor on the username and choose **My Credentials** from the dropdown list. Copy the IAM username and account name to **username** and **domainname**, and enter the IAM user password.

Figure 3-12 Token-based authentication

Q Filter	POST get-token × + ····
History Collections APIs	▶ get-token
+ New Collection Trash	
SDK Moderation	POST
2 requests	Params Authorization Headers (9) Body Pre-request Script Tests Settings
POST get-token	none form-data x-www-form-urlencoded sraw binary GraphQL JSON v
POST Moderation_image	<pre>1 * { 2 * "auth": { 3 * "identity": { 4 * "methods": [5 "password" 6], 7 * "password": { 8 * "user": { 9 "name": "username", 9 "name": "xxxxxxxxx, 11 * "domin": { 12 "name": "dominname" 13 } 14 } 15 } 16 , 17 * "scope": { 18 * "project": { 19 "name": "cn-north-4" 20 } 22 } </pre>

Click Send to send a request, and obtain and copy the token.
 In Headers, obtain the value of X-Subject-Token, which is the token. The token is valid for 24 hours.

Figure 3-13 Obtaining a token

Q Filter	POST get-token + ••••	No Environment
History Collections APIs + New Collection Trash	POST • https://iam.cn-north-4.myhuaweicloud.com	/v3/auth/tokens Send T
SDK_Moderation	Date ③	Wed, 12 May 2021 08:09:43 GMT
2 requests	Pragma 🚯	no-cache
POST get-token	Server (1)	api-gateway
POST Moderation_image	Expires (i)	Thu, 01 Jan 1970 00:00:00 GMT
	Connection (1)	Keep-Alive
	Content-Type ()	application/json; charset=UTF-8
	X-Request-Id 🔅	b55382d496e77123926fb39ae2aba715
	Cache-Control ()	no-cache, no-store, must-revalidate
	Content-Length 🗈	17953
	X-IAM-Trace-Id ()	token_cn-north-4_null_fdfb0b008d737acc352c1a36336df35f
	X-Frame-Options 🕕	SAMEORIGIN
	X-Subject-Token 🔅	MIIVsQYJKoZlhvcNAQcCoIIVojCCFZ4CAQExDTALBglghkgBZQMEAgEv
	X-XSS-Protection ()	1; mode=block;
	X-Download-Options 🔅	noopen
	X-Content-Type-Options (i)	nosniff
	Strict-Transport-Security (1)	max-age=31536000; includeSubdomains;

Step 6: Call a Service API

- 1. In the **Collections** directory in the navigation pane of Postman, click the **Moderation_image** configuration file.
- 2. Click **Headers** and copy the token value to **X-Auth-Token**.

Figure 3-14 Entering the token

Q Filter History Collections	APIs	POST get-token POST Moderation_image X + + > Moderation_image				
+ New Collection SDK_Moderation	Trash	POST • https://moderation.cn-north-4.myhuaweicloud.com/v1.0/moderation/image				
2 requests		Params Authorization Headers (10) Body Pre-request Script Tests Settings				
POST get-token		Headers 🐵 8 hidden				
POST Moderation_image		KEY VALUE	DESC			
		Content-Type application/json				
		X-Auth-Token				
		Key Value	Desc			
		Response				

- 3. Click **Body** and enter the Base64 code of the image to **image**.
 - For details about other parameters in the body, see the Image Moderation API.

Figure 3-15 Modifying the parameters

Params	Authorization	Headers (10)	Body 鱼	Pre-re	equest Script	Tests	Settings
none	form-data	x-www-form-ur	lencoded	🖲 raw	binary	GraphQL	JSON
1 - [{]							
2	"image":"Base6	4 code of the image",					
2 3	"url": "",	a an ann an a					
4 -	"categories":	1					
5	"politics"						
6	"terrorism	•					
7	"porn"						
5 6 7 8 9],						
9	"threshold" :	0.6					
10 }							

4. Click **Send** to send the request and obtain the API call results.

{



(Optional) Step 7: View the Number of API Calls

Image Moderation analyzes and identifies sensitive content (such as pornography) in images uploaded by users and returns the result.

To call Image Moderation, perform the following steps:

- **Step 1** Log in to the management console and choose **Service List > Content Moderation**.
- **Step 2** View the number of API calls. Choose **Service List > Image Moderation**, and view call statistics and call details. See **Figure 3-16**.

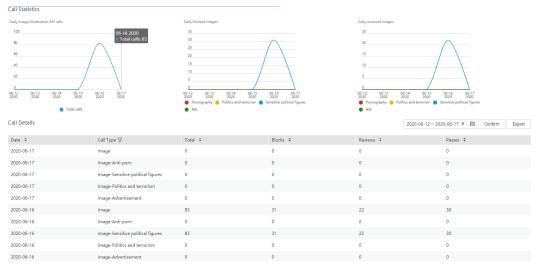


Figure 3-16 Call statistics

- The **Call Statistics** section provides three statistical charts, displaying daily calls, daily blocked text, and daily reviewed text. You can set a time range to view the change of the number of API calls within the time range.
- The **Call Details** section lists the total number of calls, blocks, reviews, and passes every day. You can click **Export** to download and view the total number of API calls, helping you better understand the service calling and review status.
- Images are stored in OBS buckets. Therefore, the images that are determined as **block** or **review** images cannot be exported.

NOTE

- **block** indicates that sensitive information is contained and the information is blocked.
- pass indicates that sensitive information is not contained and the information is approved.
- **review** indicates that manual review is required.
- **Step 3** View monitoring metrics. You can click **View Metric** to view historical data such as successful and failed calls on the Cloud Eye console. See **Figure 3-17**.

Figure 3-17 Viewing metrics

1h 3h 12h 24 hours 7d 🗐 Auto Refresh							
Period Raw data 🔹							
Successful Calls of Service ③	+ 、	Failed Calls of	f Service ⑦				+
Times/min	Max Min	Times/min					Max Min
1.2	0 0	1.2					0 0
0.9		0.9					
0.6		0.6					
0.3		0.3					
0		0					
13:50 14:02 14:14 14:26 14:38	14:50	13:50	14:02	14:14	14:26	14:38	14:50

----End

4 Using an SDK to Call Content Moderation

4.1 Moderation (Text)

This section provides an example of how to use a Java SDK to call Moderation (Text). You can directly call APIs to use SDK functions. The procedure is as follows:

Step 1: Subscribe to a Service. In the left navigation pane, choose **Service List** or **Service Management**, and subscribe to Moderation (Text).

Step 2: Configure a Custom Dictionary. Configure allowlist and blocklist dictionaries.

Step 3: Configure the Environment. Obtain the SDK and sample project and import them to the development environment.

Step 4: Perform Authentication. Use AK/SK for authentication.

Step 5: Call a Service API. Call the service API. You can view the status code and error code at any time during the usage.

Step 1: Subscribe to a Service

Subscribe to Moderation (Text) by following the instructions in **Subscribing to the Service**.

Step 2: Configure a Custom Dictionary

You can configure allowlist and blocklist dictionaries for text moderation.

1. If an allowlist dictionary is customized, the words in the allowlist dictionary will not be detected when you use text moderation.

The procedure is as follows:

Step 1 Log in to the HUAWEI CLOUD and click **Console** in the upper right corner. The **Console** page is displayed.

- **Step 2** Click **Service List** and choose **AI** > **Content Moderation**. The **Content Moderation** management console is displayed.
- Step 3 In the left navigation pane, choose Service List > Text Moderation. Click Configuration next to Custom Allowlist Dictionary. The Create Dictionary dialog box is displayed.

* Dictionary Name Enter a dictionary name	Create Dictionary ×					
* Add Keyword ③ You have entered 0 keywords.						
Enter a keyword.						
OK Cancel						

Figure 4-1 Create Dictionary

Step 4 Enter a dictionary name and the keywords in the allowlist dictionary. Click **OK**.

- **Dictionary Name**: A dictionary name contains a maximum of 40 characters. Only letters, digits, hyphens (-), and underscores (_) are allowed.
- Add Keyword: A dictionary contains a maximum of 5000 keywords. Each keyword consists of a maximum of 40 characters and ends with a carriage return.

----End

2. If a blocklist dictionary is customized, the words in the blocklist dictionary will be detected when you use Text Moderation.

The procedure is as follows:

- **Step 1** Log in to the HUAWEI CLOUD and click **Console** in the upper right corner. The **Console** page is displayed.
- **Step 2** Click **Service List** and choose **AI** > **Content Moderation**. The **Content Moderation** management console is displayed.
- Step 3 In the left navigation pane, choose Service List > Text Moderation. Click Configuration next to Custom Blocklist Dictionary. The Custom Dictionary page is displayed.
- **Step 4** Click **Create Dictionary** in the upper left corner. The **Create Dictionary** dialog box is displayed. Enter a dictionary name and the keywords in the blocklist dictionary.
 - **Dictionary Name**: A dictionary name contains a maximum of 40 characters. Only letters, digits, hyphens (-), and underscores (_) are allowed.

- Add Keyword: A dictionary contains a maximum of 5000 keywords. Each keyword consists of a maximum of 40 characters and ends with a carriage return.
- Step 5 Click OK.

NOTE

• The dictionary name cannot be modified.

----End

Step 3: Configure the Environment

You can compile code to call the API of Moderation (Text) based on the SDK. You need to configure the environment when using the SDK and calling APIs. The procedure is as follows:

- **Step 1** To obtain the SDK and related document of Content Moderation, click here.
- **Step 2** For details about how to configure the environment, see **Preparing a Java Development Environment**.
- **Step 3** Import an SDK project on Eclipse.
 - Open Eclipse and configure the correct JRE path in Windows > Preferences > Java > Installed JREs.
 - In the navigation pane on the left, choose General > Workspace. In the Text file encoding area, select Other, set the parameter value to UTF-8, click Apply, and click OK.
 - On the Eclipse menu bar, choose File > Import. In the dialog box that is displayed, choose Maven > Existing Maven Projects, click Next, click Browse, and select the local path where ais-moderation-java-sdk resides.
 - 4. Click **Finish** to import the SDK. After the SDK is imported, open the project. **Figure 4-2** shows the project directory.

Figure 4-2 Project directory

- ✓ w ais-moderation-java-sdk
 - - ✓ ♣ com.huawei.ais.demo
 - - > 🕖 ModerationClarityDemo.java
 - > D ModerationDistortionCorrectDemo.java
 - > D ModerationImageContentBatchDemo.java
 - > 🕖 ModerationImageContentBatchJobsDemo.java
 - > 🕖 ModerationImageContentDemo.java
 - > 🕖 ModerationTextContentDemo.java
 - 🗸 🚰 vcm
 - > 🔊 ModerationLongVideoAkskDemo.java
 - > 🕖 ModerationLongVideoTokenDemo.java
 - > J HttpJsonDataUtils.java
 - > 🕖 ResponseProcessUtils.java
 - > D ServiceAccessBuilder.java
 - > 🚺 TokenDemo.java
 - > 进 src/main/resources
 - > 🛋 JRE System Library [JavaSE-1.8]
 - > 🛋 Maven Dependencies
 - > 🗁 data
 - > 🗁 lib
 - > 🐎 src
 - ≽ target
 - pom.xml
 - VERSION
- ----End

Step 4: Perform Authentication

Content Moderation supports token-based and AK/SK-based authentication. This section uses AK/SK-based authentication as an example.

Step 1 Obtain an AK/SK.

The AK/SK is the access key. To obtain the AK/SK, log in to the **My Credentials** page, choose **Access Keys** in the left navigation pane, and click **Create Access Key** in the right pane.

Figure 4-3	Cleaning an ac	LESS KEY			
My Credentials	Access Keys ⑦				
API Credentials	Access keys can be downloaded on	y once after being generated. Keep them secure, char	nge them periodically, and do not share them with any	one.	
Access Keys	⊙ Create Access Key Access keys	available for creation: 2			
	Access Key ID ↓Ξ	Description ↓≡	Created ↓≡	Status ↓≡	Operation
			No data ava	ilable.	

Figure 1-2 Creating an access kow

Step 2 Configure the AK/SK in the Java SDK for AK/SK authentication.

Change the values of AK and SK of the function in the ModerationTextContentDemo.java file of the project to the obtained AK/SK. The sample code is as follows: // 1. Configure the basic information for accessing Text Moderation and generate a client connection object. AisAccess service = ServiceAccessBuilder.builder() .ak("######") // your ak .sk("######") // your sk .region("ap-southeast-1") // Content Moderation in CN-Hong Kong (ap-southeast-1) and AP-

```
Singapore (ap-southeast-3)
                                      // Timeout limit for connecting to the target URL
     .connectionTimeout(5000)
     .connectionRequestTimeout(1000)
                                        // Timeout limit for obtaining available connections from the
connection pool
     .socketTimeout(20000)
                                     // Timeout limit for obtaining server response data
     .build();
```

```
----End
```

Step 5: Call a Service API

Enter the text that you want to check in the ModerationTextContentDemo.java Step 1 file, for example, 6666666666.

```
// 2. Construct the parameters required for accessing Text Moderation.
String uri = "/v1.0/moderation/text";
JSONObject json = new JSONObject();
json.put("categories", new String[] { "porn"}) //Text for moderation
JSONObject text = new JSONObject();
text.put("text", "66666666666");
text.put("type", "content");
JSONArray items = new JSONArray();
items.add(text);
json.put("items", items);
StringEntity stringEntity = new StringEntity(json.toJSONString(), "utf-8");
// 3. Input the URI and required parameters of Text Moderation.
// Input the parameters in JSON objects and call the service using POST.
HttpResponse response = service.post(uri, stringEntity);
// 4. Check whether the API call is successful. If 200 is returned, the API call succeeds. Otherwise, it fails.
```

Step 2 Execute the ModerationTextContentDemo.java file. If 200 is displayed on the console, the program is successfully executed. The text moderation result is displayed on the console. See Figure 4-4.

ResponseProcessUtils.processResponseStatus(response);

Figure 4-4 Execution result

```
200
{
    "result": {
        "detail": {
            "flood": "6666666666"
        },
        "suggestion": "block"
    }
}
```

Step 3 View the number of API calls. Choose **Service List > Text Moderation**, and view call statistics and call details. See **Figure 4-5**.

Figure 4-5 Call statistics

Call Statistics					
Daily Text Moderation API calls		Daily blocked text		Daily reviewed text	
1		1		1	
0.8		0.8		0.8	
0.6		0.6		0.6	
0.4		0.4		0.4	
0.2		0.2		0.2	
Total calls	06-16 06-17 2020 2020	0 06-12 06-13 06-14 06-15 2020 2020 2020 2020 Total blocks	06-16 06-17 2020 2020	Total reviews	06-16 06-17 2020 2020
Call Details				2020-06-12 - 202	0-06-17 × 🗎 Confirm Export
Date \$	Call Type	Total ‡	Blocks \$	Reviews \$	Passes \$
2020-06-17	Text	0	0	0	0
2020-06-16	Text	0	0	0	0
2020-06-15	Text	0	0	0	0
2020-06-14	Text	0	0	0	0
2020-06-13	Text	0	0	0	0
2020-06-12	Text	0	0	0	0

- The **Call Statistics** section provides three statistical charts, displaying daily calls, daily blocked text, and daily reviewed text. You can set a time range to view the change of the number of API calls within the time range.
- The **Call Details** section lists the total number of calls, blocks, reviews, and passes every day. This helps you better understand service calls and their statuses.

NOTE

- block indicates that sensitive information is contained and the information is blocked.
- pass indicates that sensitive information is not contained and the information is approved.
- **review** indicates that manual review is required.
- **Step 4** View monitoring metrics. You can click **View Metric** to view historical data such as successful and failed calls on the Cloud Eye console. See **Figure 4-6**.

Figure 4-6 Viewing metrics

1h 3h 12h 24 hours 7d Image: Comparison of the second s							
Period Raw data *							
Successful Calls of Service 💿	+ 5	Failed Calls of	Service ⑦				+ 、
Times/min	Max Min	Times/min					Max Min
1.2	0 0	1.2					0 0
0.9		0.9					
0.6		0.6					
0.3		0.3					
0		0					
13:50 14:02 14:14 14:26 14:38	14:50	13:50	14:02	14:14	14:26	14:38	14:50



4.2 Moderation (Image)

This section provides an example of how to use a Java SDK to call Moderation (Image). You can directly call APIs to use SDK functions. The procedure is as follows:

Step 1: Subscribe to a Service. In the left navigation pane, choose **Service List** or **Service Management**, and subscribe to Moderation (Image).

(Optional) Step 2: Enable Object Storage Service (OBS) Authorization. If you want to use the data stored on OBS, enable OBS authorization.

Step 3: (Optional) Configure a Moderation Rule. Detect pornographic content in different scenarios.

Step 4: Configure the Environment. Obtain the SDK and sample project and import them to the development environment.

Step 5: Perform Authentication. Use AK/SK for authentication.

Step 6: Call a Service API. Call the service API. You can view the status code and error code at any time during the usage.

Step 1: Subscribe to a Service

Subscribe to Moderation (Image) by following the instructions in **Subscribing to the Service**.

(Optional) Step 2: Enable Object Storage Service (OBS) Authorization

If you want to use the data stored on OBS, enable OBS authorization. The procedure is as follows:

- 1. Log in to the management console and choose **Service List** > **Content Moderation**. In the navigation pane, choose **Service Management**.
- 2. Enable **OBS Authorization**.

NOTE

- The region of OBS must be consistent with that of Content Moderation.
- OBS authorization must be performed by an account, not by an IAM user.
- Once enabled, OBS authorization cannot be disabled.

Step 3: (Optional) Configure a Moderation Rule

You can configure review rules to detect pornographic content in different scenarios. For details, see **Configuring Moderation Rules**.

Step 4: Configure the Environment

You can compile code to call the API of Moderation (Text) based on the SDK. You need to configure the environment when using the SDK and calling APIs. The procedure is as follows:

- **Step 1** To obtain the SDK and related document of Content Moderation, click here.
- **Step 2** For details about how to configure the environment, see **Preparing a Java Development Environment**.
- **Step 3** Import an SDK project on Eclipse.
 - Open Eclipse and configure the correct JRE path in Windows > Preferences > Java > Installed JREs.
 - In the navigation pane on the left, choose General > Workspace. In the Text file encoding area, select Other, set the parameter value to UTF-8, click Apply, and click OK.
 - 3. On the Eclipse menu bar, choose **File** > **Import**. In the dialog box that is displayed, choose **Maven** > **Existing Maven Projects**, click **Next**, click **Browse**, and select the local path where **ais-moderation-java-sdk** resides.
 - 4. Click **Finish** to import the SDK. After the SDK is imported, open the project. **Figure 4-7** shows the project directory.

Figure 4-7 Project directory

- ✓ w ais-moderation-java-sdk
 - - ✓ 册 com.huawei.ais.demo
 - - > 🕖 ModerationClarityDemo.java
 - > D ModerationDistortionCorrectDemo.java
 - > D ModerationImageContentBatchDemo.java
 - > 🕖 ModerationImageContentBatchJobsDemo.java
 - > 🕖 ModerationImageContentDemo.java
 - > 🕖 ModerationTextContentDemo.java
 - 🗸 🚰 vcm
 - > 🔊 ModerationLongVideoAkskDemo.java
 - > 🕖 ModerationLongVideoTokenDemo.java
 - > J HttpJsonDataUtils.java
 - > 🕖 ResponseProcessUtils.java
 - > D ServiceAccessBuilder.java
 - > 🚺 TokenDemo.java
 - > 进 src/main/resources
 - > 🛋 JRE System Library [JavaSE-1.8]
 - > 🛋 Maven Dependencies
 - > 🗁 data
 - > 🗁 lib
 - > 🐎 src
 - *>* target
 - pom.xml
 - VERSION
- ----End

Step 5: Perform Authentication

Content Moderation supports token-based and AK/SK-based authentication. This section uses AK/SK-based authentication as an example.

Step 1 Obtain an AK/SK.

The AK/SK is the access key. To obtain the AK/SK, log in to the **My Credentials** page, choose **Access Keys** in the left navigation pane, and click **Create Access Key** in the right pane.

rigule 4-0	Cicating an ac	CC33 KCy			
My Credentials	Access Keys 🧿				
API Credentials	Access keys can be downloaded only	v once after being generated. Keep them secure, cha	nge them periodically, and do not share them with an	rone.	
Access Keys		available for creation: 2			
	Access Key ID ↓Ξ	Description ↓Ξ	Created J≡	Status ↓≡	Operation
			No data av	ailable.	

Figure 4-8 Creating an access key

Step 2 Configure the AK/SK in the Java SDK for AK/SK authentication.

Change the values of **AK** and **SK** of the function in the ModerationImageContentDemo.java file of the project to the obtained AK/SK. The sample code is as follows:

// 1. Configure the basic information for accessing Image Moderation and generate a client connection object. AisAccess service = ServiceAccessBuilder.builder() .ak("######") // your ak .sk("######") // your sk .region("ap-southeast-1") // Content Moderation in CN-Hong Kong (ap-southeast-1) and AP-Singapore (ap-southeast-3) .connectionTimeout(5000) // Timeout limit for connecting to the target URL .connectionRequestTimeout(1000) // Timeout limit for obtaining available connections from the connection pool .socketTimeout(20000) // Timeout limit for obtaining server response data .build();

----End

Step 6: Call a Service API

Step 1 Select a local image or use the default image of the sample project, and modify the image file path (**data/moderation-demo-1.jpg**) in the ModerationImageContentDemo.java file.

public static void main(String[] args) throws IOException { ModerationImageContentDemo tool = new ModerationImageContentDemo(); tool.imageContentCheck("https://sdk-obs-source-save.obs.cn-north-4.myhuaweicloud.com/XXX.jpg"); byte[] imageBytes = FileUtils.readFileToByteArray(new File("data/moderation-demo-1.jpg")); tool.imageContentCheck(imageBytes); }

Step 2 Execute the ModerationImageContentDemo.java file. If 200 is displayed on the console, the program is successfully executed. The image moderation result is displayed on the console. See Figure 4-9.

Figure 4-9 Execution result

200

{"result":{"detail":{"politics":[],"terrorism":[{"confidence":0.0,"label":"bloody"},{"confidence":0.0,"label":"fire"},{" confidence":0.0,"label":"gun"},{"confidence":0.0,"label":"knife"},{"confidence":0.0,"label":"flag"},{"confidence":0.0,"l abel":"symbol"},{"confidence":0.0,"label":"dress"},{"confidence":0.0,"label":"war"},{"confidence":0.0,"label":"tiananmen "},{"confidence":1.0,"label":"normal"}]},"suggestion":"pass"}}

Step 3 View the number of API calls. Choose Service List > Image Moderation, and view call statistics and call details. See Figure 4-10.

- 3					
Call Statistics					
Daily Image Moderation API calls		Daily blocked images		Daily reviewed images	
100	06-16 2020 Total calls: 83	35		25	
80		25	\wedge	20	\longrightarrow
60		20	/	15	/
40		15		10	
20		5		5	
0	05.15 05.16 05.17	0	06-15 06-16 06-17 2020 2020 2020	05.12 05.13 05.14	06-15 06-16 06-17 2020 2020 2020
6-12 06-13 06-14 2020 2020 2020	06-15 06-16 06-17 2020 2020 2020	06-12 06-13 06-14 2020 2020 2020 Pornography Politics and te	06-15 06-16 06-17 2020 2020 2020 rrorism Sensitive political figures	06-12 06-13 06-14 2020 2020 2020 Pornography Politics an	06-15 06-16 06-17 2020 2020 2020 id terrorism Sensitive political figures
 Tr 	stal calls	Ads		Ads	
all Details				202	20-06-12 - 2020-06-17 × 🗎 Confirm Expor
		Total \$	Blocks \$	Reviews \$	
Date ‡	Call Type 🖓				Passes \$
2020-06-17	Image	0	0	0	0
2020-06-17	Image-Anti-porn	0	0	0	0
2020-06-17	Image-Sensitive political figures	0	0	0	0
2020-06-17	Image-Politics and terrorism	0	0	0	0
2020-06-17	Image-Advertisement	0	0	0	0
2020-06-16	Image	83	31	22	30
2020-06-16	Image-Anti-porn	0	0	0	0
2020-06-16	Image-Sensitive political figures	83	31	22	30
2020-06-16	Image-Politics and terrorism	0	0	0	0
2020-06-16	Image-Advertisement	0	0	0	0

Figure 4-10 Call statistics

- The **Call Statistics** section provides three statistical charts, displaying daily calls, daily blocked text, and daily reviewed text. You can set a time range to view the change of the number of API calls within the time range.
- The **Call Details** section lists the total number of calls, blocks, reviews, and passes every day. You can click **Export** to download and view the total number of API calls, helping you better understand the service calling and review status.
- Images are stored in OBS buckets. Therefore, the images that are determined as **block** or **review** images cannot be exported.

NOTE

- block indicates that sensitive information is contained and the information is blocked.
- pass indicates that sensitive information is not contained and the information is approved.
- **review** indicates that manual review is required.
- **Step 4** View monitoring metrics. You can click **View Metric** to view historical data such as successful and failed calls on the Cloud Eye console. See **Figure 4-11**.

Figure 4-11 Viewing metrics

1h 3h 12h 24 hours 7d Image: The second s	D Refresh	
Successful Calls of Service 💿	Failed Calls of Service ③	+ 5
Times/min	Max Min Times/min	Max Min
1.2	0 0 1.2	0 0
0.9	0.9	
0.6	0.6	
0.3	0.3	
0	0	
13:50 14:02 14:14 14:26	14:38 14:50 13:50 14:02 14:14 14:26 14:38	14:50

