

ALIBABA CLOUD

# Alibaba Cloud

内容安全  
User Guide

Document Version: 20220524

 Alibaba Cloud

# Legal disclaimer

Alibaba Cloud reminds you to carefully read and fully understand the terms and conditions of this legal disclaimer before you read or use this document. If you have read or used this document, it shall be deemed as your total acceptance of this legal disclaimer.

1. You shall download and obtain this document from the Alibaba Cloud website or other Alibaba Cloud-authorized channels, and use this document for your own legal business activities only. The content of this document is considered confidential information of Alibaba Cloud. You shall strictly abide by the confidentiality obligations. No part of this document shall be disclosed or provided to any third party for use without the prior written consent of Alibaba Cloud.
2. No part of this document shall be excerpted, translated, reproduced, transmitted, or disseminated by any organization, company or individual in any form or by any means without the prior written consent of Alibaba Cloud.
3. The content of this document may be changed because of product version upgrade, adjustment, or other reasons. Alibaba Cloud reserves the right to modify the content of this document without notice and an updated version of this document will be released through Alibaba Cloud-authorized channels from time to time. You should pay attention to the version changes of this document as they occur and download and obtain the most up-to-date version of this document from Alibaba Cloud-authorized channels.
4. This document serves only as a reference guide for your use of Alibaba Cloud products and services. Alibaba Cloud provides this document based on the "status quo", "being defective", and "existing functions" of its products and services. Alibaba Cloud makes every effort to provide relevant operational guidance based on existing technologies. However, Alibaba Cloud hereby makes a clear statement that it in no way guarantees the accuracy, integrity, applicability, and reliability of the content of this document, either explicitly or implicitly. Alibaba Cloud shall not take legal responsibility for any errors or lost profits incurred by any organization, company, or individual arising from download, use, or trust in this document. Alibaba Cloud shall not, under any circumstances, take responsibility for any indirect, consequential, punitive, contingent, special, or punitive damages, including lost profits arising from the use or trust in this document (even if Alibaba Cloud has been notified of the possibility of such a loss).
5. By law, all the contents in Alibaba Cloud documents, including but not limited to pictures, architecture design, page layout, and text description, are intellectual property of Alibaba Cloud and/or its affiliates. This intellectual property includes, but is not limited to, trademark rights, patent rights, copyrights, and trade secrets. No part of this document shall be used, modified, reproduced, publicly transmitted, changed, disseminated, distributed, or published without the prior written consent of Alibaba Cloud and/or its affiliates. The names owned by Alibaba Cloud shall not be used, published, or reproduced for marketing, advertising, promotion, or other purposes without the prior written consent of Alibaba Cloud. The names owned by Alibaba Cloud include, but are not limited to, "Alibaba Cloud", "Aliyun", "HiChina", and other brands of Alibaba Cloud and/or its affiliates, which appear separately or in combination, as well as the auxiliary signs and patterns of the preceding brands, or anything similar to the company names, trade names, trademarks, product or service names, domain names, patterns, logos, marks, signs, or special descriptions that third parties identify as Alibaba Cloud and/or its affiliates.
6. Please directly contact Alibaba Cloud for any errors of this document.

# Document conventions

Style	Description	Example
 <b>Danger</b>	A danger notice indicates a situation that will cause major system changes, faults, physical injuries, and other adverse results.	 <b>Danger:</b> Resetting will result in the loss of user configuration data.
 <b>Warning</b>	A warning notice indicates a situation that may cause major system changes, faults, physical injuries, and other adverse results.	 <b>Warning:</b> Restarting will cause business interruption. About 10 minutes are required to restart an instance.
 <b>Notice</b>	A caution notice indicates warning information, supplementary instructions, and other content that the user must understand.	 <b>Notice:</b> If the weight is set to 0, the server no longer receives new requests.
 <b>Note</b>	A note indicates supplemental instructions, best practices, tips, and other content.	 <b>Note:</b> You can use Ctrl + A to select all files.
>	Closing angle brackets are used to indicate a multi-level menu cascade.	Click <b>Settings &gt; Network &gt; Set network type</b> .
<b>Bold</b>	Bold formatting is used for buttons, menus, page names, and other UI elements.	Click <b>OK</b> .
Courier font	Courier font is used for commands	Run the <code>cd /d C:/window</code> command to enter the Windows system folder.
<i>Italic</i>	Italic formatting is used for parameters and variables.	<code>bae log list --instanceid</code> <i>Instance_ID</i>
[ ] or [a b]	This format is used for an optional value, where only one item can be selected.	<code>ipconfig [-all -t]</code>
{ } or {a b}	This format is used for a required value, where only one item can be selected.	<code>switch {active stand}</code>

# Table of Contents

- 1.Content Moderation API ----- 05
  - 1.1. Customize policies for machine-assisted moderation ----- 05
  - 1.2. Enable callback notifications ----- 10
  - 1.3. Manage sample libraries ----- 15
    - 1.3.1. Manage custom image libraries ----- 15
    - 1.3.2. Manage custom text libraries ----- 18
  - 1.4. Review machine-assisted moderation results ----- 24
  - 1.5. View statistics ----- 26
  - 1.6. Give feedback on samples ----- 29
  - 1.7. Authorize a role to access ApsaraVideo for Media Processin...----- 29
- 2.RAM user permissions ----- 31
  - 2.1. Grant read-only permissions of Content Moderation to a R... ----- 31
  - 2.2. Call the Content Moderation API as a RAM user ----- 33

# 1.Content Moderation API

## 1.1. Customize policies for machine-assisted moderation

Content Moderation uses the default moderation policy of Alibaba Cloud for machine-assisted moderation. If you find that the default moderation policy is too strict or loose and cannot meet your business requirements, you can customize policies for machine-assisted moderation in Content Moderation. This topic describes how to customize a policy for machine-assisted moderation.

### Context

A BizType setting specifies a business scenario. Moderation policies are customized based on the business scenario. Each business scenario corresponds to a moderation policy. If you do not customize moderation policies, the default business scenario and corresponding default moderation policy are used. After you customize a policy for machine-assisted moderation based on a business scenario, you can specify the business scenario in an API request for content moderation. In this case, the corresponding moderation policy takes effect.

### Procedure

1. Log on to the [Content Moderation console](#).
2. In the left-side navigation pane, choose **Settings > Machine audit**.
3. Create a business scenario.

If you have created a business scenario, skip this step.

- i. On the **BizType Manage** tab of the **Machine audit** page, click **Create BizType**.

- ii. In the **Create bizType** dialog box, set the parameters that are described in the following table. You can set a name and specify an industry template and the industry classification for the business scenario.

Parameter	Description
Name	The name of the business scenario. The name can contain digits, letters, and underscores (_). It can be up to 32 characters in length.
CiteTemplate	Specifies whether to import the configuration of an industry template. If you use an industry template, specific features of the moderation policy cannot be customized. If you do not use an industry template, you can customize the moderation policy.
Industry	The industry classification of the business scenario. An appropriate industry classification allows Alibaba Cloud to help you adjust your moderation policy.
Import	If you have created a business scenario, you can import the configuration of the existing business scenario.
Description	The description of the business scenario. The description can contain letters, digits, and underscores (_). It can be up to 32 characters in length.

- iii. Click **OK**. The business scenario is created.

4. Configure the business scenario.

To configure the business scenario, complete configurations on the **Moderation Policy**, **Audit Data**, and **Evidence Storage** tabs.

- o Configurations on the **Moderation Policy** tab

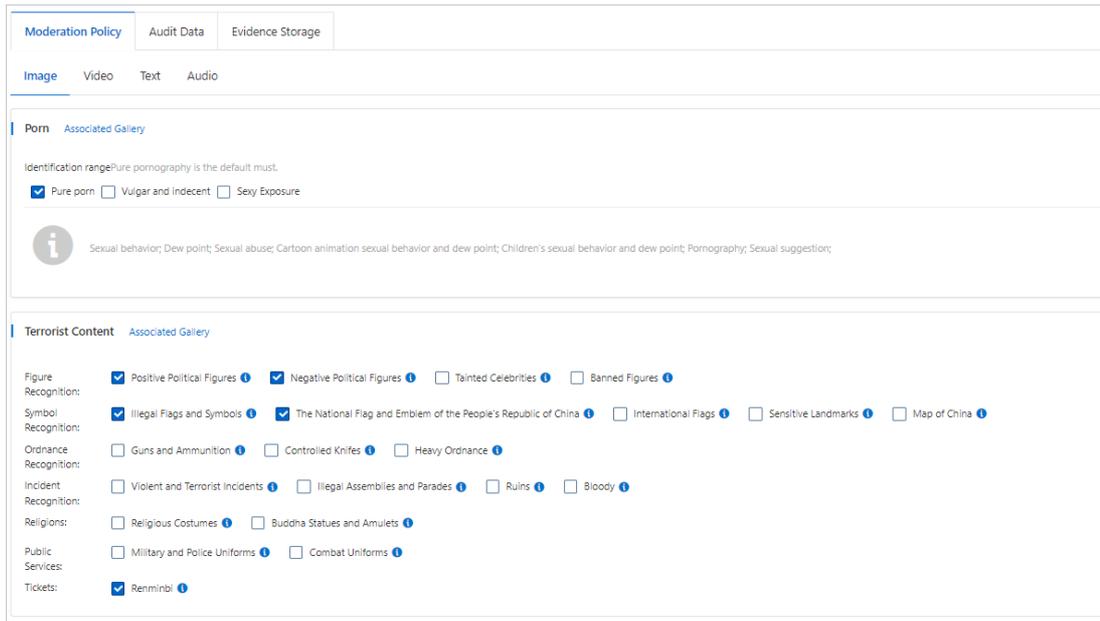
The moderation policy template varies based on the moderation scenario. For more information, see the templates in the console.

- a. On the **Machine audit** page, find the business scenario that you want to configure and click **Edit** in the Actions column.

b. Customize the policy for machine-assisted moderation based on your business requirements.

Content Moderation allows you to customize the moderation policy for images, videos, text, and audio. For images and videos, machine-assisted moderation covers pornography detection, terrorist content detection, undesirable scene detection, and ad violation detection. For text and audio, machine-assisted moderation detects pornographic content, political content, abuse, ads, and prohibited content.

If a specific detection scenario is not required for your business, do not select the detection scenario. For example, if you do not want to moderate ads, you can clear all ad-related detection scenarios when you customize the policy for machine-assisted moderation.



**Note** When you customize the moderation policy for images, you can click **Associated Gallery** or **Associated Text Library** in the upper-right corner of a section to associate the moderation policy with the configurations of image libraries or text libraries.

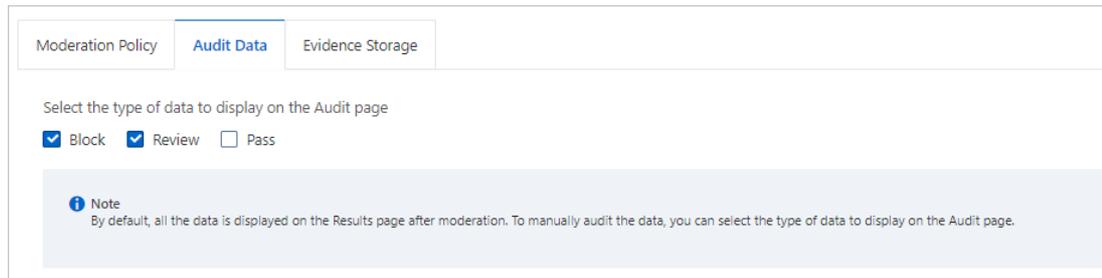
c. Click **Save**.

If you need to use the customized policy for machine-assisted moderation when you call an API operation, you must specify the business scenario that corresponds to the moderation policy. For example, when you call an API operation to moderate images for pornographic content, set the scenes parameter to *porn* in the API request. In this case, the corresponding moderation policy takes effect. For more information, see [Synchronous moderation](#).

o Configurations on the **Audit Data** tab

a. Click the **Audit Data** tab.

- b. On the **Audit Data** tab, select the type of data to be reviewed.



The screenshot shows a web interface with three tabs: 'Moderation Policy', 'Audit Data' (which is selected and highlighted in blue), and 'Evidence Storage'. Below the tabs, there is a heading 'Select the type of data to display on the Audit page'. Underneath, there are three radio button options: 'Block' (checked), 'Review' (checked), and 'Pass' (unchecked). Below the radio buttons is a light blue note box with an information icon and the text: 'Note: By default, all the data is displayed on the Results page after moderation. To manually audit the data, you can select the type of data to display on the Audit page.'

For more information about the human review feature, see [Review machine-assisted moderation results](#).

- o Configurations on the **Evidence Storage** tab

When you call Content Moderation operations to moderate content, you can enable the feature of storing evidence in Object Storage Service (OSS) buckets to retain the evidence that is obtained from machine-assisted moderation results. The evidence storage feature must be used with OSS. You can use this feature to store videos, audio, and images in which illegal, suspicious, or normal content is identified in specified OSS buckets. Corresponding OSS URLs are generated for the stored content. This section describes how to enable and configure the feature of storing evidence in OSS buckets.

You can store only videos, audio, and images in which illegal, suspicious, or normal content is identified in specified OSS buckets. The stored content refers to the videos, audio, and images whose machine-assisted moderation result returned in the suggestion parameter is block, review, or pass. A value of block indicates illegal, a value of review indicates suspicious, and a value of pass indicates normal.

By default, the evidence storage feature is disabled. To use this feature, you must enable it and configure settings for storing videos, audio, and images in which illegal, suspicious, or normal content is identified in OSS buckets.

- If the feature of storing video evidence in a specified OSS bucket is enabled, video files or streams whose machine-assisted moderation result is block, review, or pass and the related video snapshots are stored in the specified OSS bucket.
- If the feature of storing audio evidence in a specified OSS bucket is enabled, audio files or streams whose machine-assisted moderation result is block, review, or pass and the related audio fragments are stored in the specified OSS bucket.
- If the feature of storing image evidence in a specified OSS bucket is enabled, images whose machine-assisted moderation result is block, review, or pass are stored in the specified OSS bucket.

- a. Click the **Evidence Storage** tab.

- b. On the **Evidence Storage** tab, turn on **Enable Evidence Storage** in the **Video Evidence Storage**, **Audio Evidence Storage**, and **Image evidence transfer** sections as needed and complete required configurations.

**Note** If you log on to the Content Moderation console for the first time, you must grant the read and write permissions on OSS to the current account as prompted. You can reuse the permission policy for detecting illegal content in an OSS bucket. If the current account is granted the required permissions, the Evidence Storage tab appears.

Parameter	Description
<b>OSS Bucket</b>	The OSS bucket that is used to store the evidence files.
<b>Storage Directory</b>	<p>The folder used to store the evidence files in the specified OSS bucket. All evidence files are stored in the specified folder based on the storage rules. For more information, see <a href="#">Storage rules</a>.</p> <p><b>Note</b> If the specified folder does not exist in the OSS bucket, the system automatically creates one.</p>
<b>URL Validity Duration</b>	The validity period of the OSS URL of each evidence file that is stored in the OSS bucket. Valid values: 300 to 3600. Unit: seconds.
<b>Transfer range</b>	<p>Valid values:</p> <ul style="list-style-type: none"> <li><b>Machine identification violation:</b> stores the content whose machine-assisted moderation result is block.</li> <li><b>Machine identification is suspected:</b> stores the content whose machine-assisted moderation result is review.</li> <li><b>The machine identification is normal.:</b> stores the content whose machine-assisted moderation result is pass.</li> </ul>

- c. Click **Save**.

After you complete the configurations for the evidence storage feature, the detected illegal, suspicious, and normal content is stored in the specified OSS buckets when you call API operations to review videos and images and detect spam in audio. You can view the OSS URLs of the stored video-related files in the returned `data.extras.newUrl` parameter, the OSS URLs of the stored audio-related files in the returned `data.new_url` and `data.result.details.url` parameters, and the OSS URLs of the stored image files in the returned `data.storedUrl` parameter. You can also log on to the [OSS console](#) to view the files stored in the specified buckets.

### Storage rules

Storage rules for video-related files

File type	Storage path	Naming convention
Video file	<code>/\${bucket}/\${Storage folder}/video/\${suggestion}/\${taskId}/\${Video name.Suffix}</code>	A video file stored in the OSS bucket is named after the original video file.

File type	Storage path	Naming convention
Video snapshot	$\${bucket}/\${Storage\ folder}/video/\${suggestion}/\${taskId}/\${Video\ snapshot\ name.Suffix}$	A video snapshot file stored in the OSS bucket is named based on the point in time when the snapshot is captured. For example, <i>00_01_02</i> indicates that the snapshot was captured at 00:01:02.
Video stream	$\${bucket}/\${Storage\ folder}/video/\${suggestion}/\${taskId}/\${Video\ stream\ name.Suffix}$	A video stream file stored in the OSS bucket is named based on the point in time when the review starts. For example, the name <i>20190102_12_02_03.wav</i> indicates that the review of the video stream started at 12:02:03 on January 2, 2019.

#### Storage rules for audio-related files

File type	Storage path	Naming convention
Audio file	$\${bucket}/\${Storage\ folder}/audio/\${suggestion}/\${taskId}/\${Audio\ name.Suffix}$	An audio file stored in the OSS bucket is named after the original audio file.
Audio fragment	$\${bucket}/\${Storage\ folder}/audio/\${suggestion}/\${taskId}/\${Audio\ fragment\ name.Suffix}$	An audio fragment file stored in the OSS bucket is named based on the start and end points in time of the audio fragment. For example, the name <i>00_01_02-00_10_13.mp3</i> indicates that the audio fragment is trimmed from 00:01:02 to 00:10:13 of the original audio file.

#### Storage rules for image files

File type	Storage path	Naming convention
Image	$\${bucket}/\${Storage\ folder}/image/\${suggestion}/\${taskId}/\${Image\ name.Suffix}$	An image file stored in the OSS bucket is named after the original image file.

## 1.2. Enable callback notifications

Content Moderation can send asynchronous notifications to inform you of machine-assisted moderation results and your review results. If you want to use or integrate the results for your business purpose, you can enable callback notifications. This topic describes how to enable callback notifications for the Content Moderation API.

### Context

The Content Moderation API supports callback notifications for machine-assisted moderation results and human review results.

- Callback notification for machine-assisted moderation results: After a moderation request is processed, Content Moderation sends the machine-assisted moderation results to the specified HTTP callback URL by

sending an HTTP POST request.

- **Callback notification for human review results:** After you review data or call feedback operations to rectify the machine-assisted moderation results, Content Moderation sends the human review results to the specified HTTP callback URL by sending an HTTP POST request. For more information, see [Review machine-assisted moderation results](#).

## Terms

The following table introduces the concepts related to callback notifications.

Concept	Description
callback URL	<p>The public endpoint of your server that you specify in the <a href="#">Content Moderation console</a>. The callback URL must meet the following requirements:</p> <ul style="list-style-type: none"> <li>• Uses HTTP or HTTPS and is accessible from the Internet.</li> <li>• Supports the POST method.</li> <li>• Supports UTF-8-encoded data.</li> <li>• Supports data that is received in the <code>application / x-www-form-urlencoded</code> format.</li> <li>• Supports the checksum and content parameters in callback data.</li> </ul>
seed	The string that is used to verify whether the POST request is sent from Content Moderation to the specified HTTP callback URL.
callback times	The number of times that Content Moderation pushes callback notifications to your server. If your server receives a callback notification, it sends the HTTP status code 200 to Content Moderation. If your server fails to receive a callback notification, it sends other HTTP status codes to Content Moderation. After Content Moderation receives an HTTP status code other than 200, Content Moderation continues to push the callback notification until your server receives it. Content Moderation can repeatedly push a callback notification for up to 16 times.
callback data	The content of the callback notification that Content Moderation sends to the specified callback URL. For more information about the parameters in the callback data, see <a href="#">Parameters in the callback data</a> .

### Parameters in the callback data

Parameter	Type	Description
checksum	String	<p>The string in the <code>&lt;UID&gt; + &lt;Seed&gt; + &lt;Content&gt;</code> format that is generated by the Secure Hash Algorithm 256 (SHA-256) algorithm. UID indicates the ID of your Alibaba Cloud account. You can query the ID in the <a href="#">Alibaba Cloud Management Console</a>.</p> <div style="background-color: #e6f2ff; padding: 5px; border: 1px solid #d9e1f2;"> <p> <b>Note</b> To prevent data tampering, you can use the SHA-256 algorithm to generate a string when your server receives a callback notification and verify the string against the received checksum parameter.</p> </div>

Parameter	Type	Description
content	String	The JSON-formatted string to be parsed to the callback data in the JSON format. For more information about the callback data that is parsed from the content parameter, see the "Content parameter description" section of this topic.

## Callback notification for machine-assisted moderation results

All asynchronous moderation operations of the Content Moderation API support callback notifications, including asynchronous image moderation and asynchronous video moderation. For more information, see [Asynchronous image moderation](#) and [Asynchronous video moderation](#). If you call an asynchronous operation and need Content Moderation to return the moderation results, set the callback and seed parameters in your moderation request. The callback parameter specifies the callback URL and the seed parameter specifies a string that is used to verify the callback notification request.

 **Note** If you do not set the preceding parameters when you call an asynchronous operation, you can only poll the asynchronous moderation results at regular intervals.

### Procedure

1. Prepare the HTTP callback URL and verification string that are used to receive moderation results.
2. When you call an asynchronous operation for content moderation, set the callback and seed parameters in your moderation request. For more information, see the parameter description of the relevant API operation.

## Callback notification for human review results

If you send human review requests, the review results are not returned in real time. You must configure callback notifications to receive human review results. Content Moderation allows you to use both the machine-assisted moderation and human review services or use only the human review service. When you configure callback notifications, take note of the following differences.

- **Use both the machine-assisted moderation and human review services**

If you use both the machine-assisted moderation and human review services, you must configure a callback notification for human review results in the Content Moderation console. Perform the following steps:

- i. Log on to the [Content Moderation console](#).
- ii. In the left-side navigation pane, choose **Settings > Machine audit**.
- iii. On the **Machine audit** page, click the **Message** tab.
- iv. On the **Message** tab, click **New notification scheme**.
- v. In the **New notification scheme** dialog box, set the **Scheme name**, **Callback URL**, **Encryption algorithm**, **Notification type**, and **Audit Result** parameters. Click **OK**.

After you configure the callback notification, the system automatically generates a value for the seed parameter. You can use the value of the seed parameter to verify whether a callback notification request is sent from Alibaba Cloud. Save the generated value of the seed parameter.

### Notice

- If you have configured a callback notification for machine-assisted moderation results, you can reuse the configurations of the callback notification or configure another callback notification based on your business requirements.
- You can configure a callback notification for receiving all moderation results or only specified moderation results based on your business scenario.
- Select **Alibaba Cloud Human Review Results** for the Notification type parameter.
- Valid values of the Encryption algorithm parameter:
  - **SHA256**: The HMAC-SHA256 encryption algorithm is used.
  - **National secret sm3**: The HMAC-SM3 encryption algorithm is used and a hexadecimal string is returned. The string consists of lowercase letters and digits.

For example,

`66c7f0f462eeedd9d1f2d46bdc10e4e24167c4875cf2f7a2297da02b8f4ba8e0` is returned after you encrypt `abc` by using the HMAC-SM3 encryption algorithm.

- vi. On the **Machine audit** page, click the **BizType Manage** tab.
- vii. On the **BizType Manage** tab, find the business scenario that you want to manage and click **choose message** in the **Actions** column to associate the **Callback notification scheme** that you configure with the business scenario.

## Content parameter description

After you enable callback notifications, Content Moderation sends the moderation results that are generated by the Content Moderation API in a callback notification to the specified callback URL. The callback notification contains the content parameter. The following tables describe the structure of the content parameter.

Structure of the content parameter

Parameter	Type	Required	Description
scanResult	JSONObject	No	<p>The machine-assisted moderation result. The structure of this parameter varies depending on the moderation object, such as images and videos.</p> <ul style="list-style-type: none"> <li>• For images, the structure is the same as that of the results parameter that is returned in response to synchronous operations for image moderation. For more information, see <a href="#">Synchronous image moderation</a>.</li> <li>• For videos, the structure is the same as that of the results parameter that is returned in response to asynchronous operations for video moderation. For more information, see <a href="#">Asynchronous video moderation</a>.</li> </ul>

Parameter	Type	Required	Description
auditResult	JSONObject	No	<p>The human review result that is generated by you. This parameter is returned only when human review is performed. For more information, see <a href="#">auditResult</a>.</p> <div style="border: 1px solid #ccc; background-color: #e6f2ff; padding: 5px; margin-top: 10px;"> <p> <b>Note</b> This parameter is not returned if Content Moderation sends only machine-assisted moderation results.</p> </div>
humanAuditResult	JSONObject	No	<p>The human review result that is generated by the human review service of Alibaba Cloud. If you purchased the human review service of Alibaba Cloud, this parameter is returned to send the human review result. For more information, see <a href="#">humanAuditResult</a>.</p>

#### auditResult

Parameter	Type	Required	Description
suggestion	String	Yes	<p>The suggestion that you provided during human review. Valid values:</p> <ul style="list-style-type: none"> <li>• <i>block</i></li> <li>• <i>pass</i></li> </ul>
labels	JSONArray	No	<p>The tag that you set during human review. The return value can be one or more tags. Valid values:</p> <ul style="list-style-type: none"> <li>• <i>porn</i></li> <li>• <i>terrorism</i></li> <li>• <i>ad</i></li> <li>• <i>live</i></li> </ul>

#### humanAuditResult

Parameter	Type	Required	Description
suggestion	String	Yes	<p>The suggestion that is provided by the human review service of Alibaba Cloud. Valid values:</p> <ul style="list-style-type: none"> <li>• <i>block</i></li> <li>• <i>pass</i></li> </ul>
taskId	String	Yes	<p>The ID of the moderation task. You can associate the human review result of the content with the corresponding machine-assisted moderation result based on the task ID.</p>
dataId	String	Yes	<p>The ID of the moderated content.</p>

Parameter	Type	Required	Description
labels	StringArray	No	<p>The tag that the human review service sets. The return value can contain multiple tags.</p> <p><b>Note</b> By default, this parameter is not returned. In addition, you must pay a fee for this parameter. For more information, contact Alibaba Cloud technical support.</p>

#### Example of the content parameter

```
{
  "scanResult": {
    "code": 200,
    "msg": "OK",
    "taskId": "fdd25f95-4892-4d6b-aca9-7939bc6e9baa-1486198766695",
    "url": "http://1.jpg",
    "results": [
      {
        "rate": 100,
        "scene": "porn",
        "suggestion": "block",
        "label": "porn"
      }
    ]
  },
  "auditResult": {
    "suggestion": "block",
    "labels": [
      "porn",
      "ad",
      "terrorism"
    ]
  },
  "humanAuditResult": {
    "suggestion": "pass",
    "dataId": "yyyy",
    "labels": [
      "Pornographic content",
      "Vulgar content"
    ]
  },
  "taskId": "xxxxxxx"
}
```

## 1.3. Manage sample libraries

### 1.3.1. Manage custom image libraries

By default, Content Moderation provides the moderation service by using the global risky image library of Alibaba Cloud. This library can meet most moderation needs. Content Moderation supports custom image libraries. You can use custom image libraries to ensure that moderation results meet specific business needs. You can specify the images to be blocked, passed, or reviewed in different custom image libraries to meet specific management needs.

## Context

Custom image libraries can be divided into blacklists, whitelists, and review lists based on the purposes. If custom image libraries are used for moderation, the images to be moderated that hit samples in custom image libraries are tagged with the corresponding category of moderation results. The moderation result for images that hit samples in blacklists is block. The moderation result for images that hit samples in whitelists is pass. The moderation result for images that hit samples in review lists is review. Custom image libraries include feedback-based image libraries and self-managed image libraries:

- Feedback-based image libraries are automatically created to accommodate the images that are reviewed. By default, you can use feedback-based image libraries to moderate images in all moderation scenarios of the same type. You can manage the images in feedback-based image libraries. However, you cannot perform operations on feedback-based image libraries. For example, you cannot disable or delete a feedback-based image library.
- Self-managed image libraries are libraries that you create to moderate images in a specific moderation scenario or a specific type of moderation scenario. You can manage the images in self-managed image libraries and perform operations on self-managed image libraries.

 **Note** You can create up to 10 self-managed image libraries and add up to 10,000 images to each self-managed image library.

When you use the Content Moderation API, you can use custom image libraries to moderate images or videos for pornographic content, terrorist content, ad violations, and undesirable scenes.

This topic describes how to manage custom image libraries for the Content Moderation API in the Alibaba Cloud Content Moderation console. You can also manage custom image libraries by calling API operations or using Content Moderation SDKs. For more information, see the following topics:

- [Call API operations to manage custom image libraries](#)
- [Use Content Moderation SDK for Java to manage custom image libraries](#)

## Procedure

1. Log on to the [Content Moderation console](#).
2. In the left-side navigation pane, choose **Machine audit > Risk Libraries**.
3. On the **Risk Libraries** page, click the Custom Image Library tab and then click **Create Image Library**.
4. In the **Create Image Library** dialog box, set the parameters and click **OK**.

Parameters for creating an image library

Parameter	Description
<b>Name</b>	The name of the custom image library. The name can be up to 64 characters in length. We recommend that you give the image library a meaningful name.

Parameter	Description
Scene	<p>The moderation scenario to which the custom image library applies. Valid values:</p> <ul style="list-style-type: none"> <li>◦ <b>Porn</b>: moderates images or videos for pornographic content, where the value of the scenes parameter contains porn in API requests.</li> <li>◦ <b>Terrorism</b>: moderates images or videos for terrorist content, where the value of the scenes parameter contains terrorism in API requests.</li> <li>◦ <b>Ad</b>: moderates images or videos for ad violations, where the value of the scenes parameter contains ad in API requests.</li> <li>◦ <b>Undesirable Scene</b>: moderates images or videos for undesirable scenes, where the value of the scenes parameter contains live in API requests.</li> </ul>
List Category	<p>The category of the moderation result that is returned based on the custom image library. Valid values:</p> <ul style="list-style-type: none"> <li>◦ <b>Blacklist</b>: If the images to be moderated hit samples in the image library, the machine-assisted moderation result contains the suggestion parameter that has a value of block.</li> <li>◦ <b>Review List</b>: If the images to be moderated hit samples in the image library, the machine-assisted moderation result contains the suggestion parameter that has a value of review.</li> <li>◦ <b>Whitelist</b>: If the images to be moderated hit samples in the image library, the machine-assisted moderation result contains the suggestion parameter that has a value of pass.</li> </ul>
bizType	<p>The business scenario to which the custom image library applies. You can specify different image libraries in API requests to meet your business needs. For example, you can use the bizType parameter to specify the image library to be applied in a specific moderation scenario. The bizType parameter takes effect in the following ways:</p> <ul style="list-style-type: none"> <li>◦ If the bizType parameter in a moderation request is set to A, the image libraries of which the bizType parameter is set to A are used for moderation. These image libraries can be used only if they are enabled.</li> <li>◦ In other cases, all enabled image libraries are used for moderation.</li> </ul>

The custom image library is created. The created image library is enabled by default.

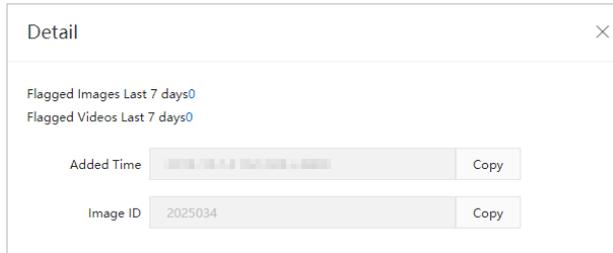
5. On the **Custom Image Library** tab, find the created image library and click **Manage** in the Actions column.

The **Custom Image Library** tab lists all custom image libraries. The libraries marked with **System** and named in the SCENARIO\_FEEDBACK\_WHITE or SCENARIO\_FEEDBACK\_BLACK format are feedback-based image libraries. For example, the LIVE\_FEEDBACK\_BLACK library is a blacklist that consists of samples added by the system and is used to moderate images for undesirable scenes.

6. On the **Image Libraries** page, perform the following operations to manage images as needed.

 **Note** You can add and delete images. The operations take effect in about 15 minutes.

- Query images by setting the **Image ID** and **Time** parameters.
- Click an image. In the **Details** panel, view the number of images or videos that hit the image, the time when the image was added, and the ID of the image.



- Click **Select File** and upload images to the current image library.

**Note** You can upload images in the *PNG, JPG, JPEG, and BMP* formats.

- Click **Delete** for an image to delete the image. Alternatively, select multiple images and click **Batch Delete** in the lower part of the page.

## Delete, modify, or disable an image library

On the **Custom Image Library** tab, you can click **Delete**, **Edit**, or **Disable** in the **Actions** column to delete, modify, or disable a self-managed image library.

## 1.3.2. Manage custom text libraries

Content Moderation supports custom text libraries. You can use custom text libraries to ensure that moderation results meet specific business needs. You can use custom text libraries for text violation detection in images, ad violation detection, text anti-spam, file anti-spam, and audio anti-spam. You can specify the text to be blocked, passed, or reviewed in different custom text libraries to meet specific management needs.

### Context

**Notice** We recommend that you follow the instructions in this topic to use custom text libraries. This prevents you from adding improper terms that affect the accuracy of moderation results.

Custom text libraries include feedback-based text libraries and self-managed text libraries:

- Feedback-based text libraries are automatically created to accommodate the text that is reviewed. By default, you can use feedback-based text libraries to moderate text in all moderation scenarios of the same type. You can manage the text in feedback-based text libraries. However, you cannot perform operations on feedback-based text libraries. For example, you cannot disable or delete a feedback-based text library. For more information about human review, see [Review machine-assisted moderation results](#).
- Self-managed text libraries are libraries that you create to moderate text in a specific moderation scenario or a specific type of moderation scenario. You can manage the text in self-managed text libraries and perform operations on self-managed text libraries.

**Note** You can create up to 10 self-managed text libraries.

When you use the Content Moderation API, you can use custom text libraries to detect text violation in images and implement text anti-spam.

This topic describes how to manage custom text libraries for the Content Moderation API in the Alibaba Cloud Content Moderation console. You can also manage custom text libraries by calling API operations or using Content Moderation SDKs. For more information, see the following topics:

- [Call API operations to manage custom text libraries](#)

- [Use Content Moderation SDK for Java to manage custom text libraries](#)

## Text types

You can add terms and text patterns to custom text libraries.

- Terms

Terms are designed to moderate words in text. If a sentence or a piece of text contains a specific term, the term is hit. You can add different terms for different business scenarios.

In Content Moderation, you can apply term-based moderation to text violation detection in images and text anti-spam. For more information about relevant parameters, see the parameter description of moderation operations in different scenarios. The relevant parameters in these two scenarios may be slightly different.

You can add the AND (&) and NOT (~) logical operators in Chinese terms. Examples:

- The term "A&B" is added. If a piece of text contains both A and B, the term is hit.
- The term "A~B" is added. If a piece of text contains A but does not contain B, the term is hit.

 **Note** If you add both logical operators in a term, the AND (&) operator must be added before the NOT (~) operator. For example, you can add "A&B~C" as a term, but cannot add "A~C&B" as a term.

- Text patterns

Text patterns are designed to compare the similarity between sentences or pieces of text. If two sentences or two pieces of text are partially different but express the same meaning, the two sentences or two pieces of text show a close similarity. Content Moderation can determine whether a piece of text has a close similarity to a text pattern in text pattern libraries. If the similarity reaches a specific degree, the text pattern is hit.

You can apply text pattern libraries to text anti-spam. Content Moderation allows you to customize a blacklist, a whitelist, and a review list for text pattern libraries based on your business needs. The review list contains the text that needs human review. You can manage text patterns related to your business in text pattern libraries. In this case, the content that hits text patterns can be filtered out in text anti-spam.

## Limits

Type	Item	Limit
Self-managed text library	Quantity	Supports a maximum of 10 self-managed text libraries.
Self-managed text library	Name length	Supports a maximum of 20 characters in length for each library name.
Term	Term type	<ul style="list-style-type: none"> <li>• Supports Chinese terms.</li> <li>• Supports letters and digits as terms.</li> </ul> <p> <b>Note</b> Each combination of letters and digits is considered as a word during word-breaking.</p> <ul style="list-style-type: none"> <li>• English words or phrases cannot be used as terms.</li> </ul>

Type	Item	Limit
Term	Number of terms in a text library	Supports a maximum of 10,000 terms in a text library.
Term	Term length	Supports a maximum of 50 characters in length for each term, including logical operators.
Term	Encoding for Chinese terms	UTF-8
Term	Term format	<p>The following special characters in full-width and half-width forms are not supported:</p> <p>At signs (@), number signs (#), dollar signs (\$), percent signs (%), carets (^), asterisks (*), parentheses (()), angle brackets (&lt;&gt;), backslashes (/), question marks (?), commas (,), periods (.), semicolons (;), underscores (_), plus signs (+), hyphens (-), equal signs (=), single quotation marks ('), double quotation marks ("), spaces, and tabs.</p>
Text pattern	Text pattern length	<p>Supports 10 to 4,000 characters in length for each text pattern.</p> <div style="background-color: #e1f5fe; padding: 5px;"> <p> <b>Note</b> If the text added to the text library is too long, invalid matches may occur. We recommend that you set each text pattern to a maximum of 200 characters in length.</p> </div>
Text pattern	Number of text patterns in a text library	Supports a maximum of 10,000 text patterns in a text library.
Text pattern	Encoding	UTF-8
Text pattern	Text content	<p>Requires clear Chinese semantic traits that can be extracted. If few semantic traits can be identified from a text pattern, this text pattern is ignored.</p> <div style="background-color: #e1f5fe; padding: 5px;"> <p> <b>Note</b> A text pattern that consists of meaningless letters, digits, or emoticons may be ignored.</p> </div>

## Procedure

1. Log on to the [Content Moderation console](#).
2. In the left-side navigation pane, choose **Machine audit > Risk Libraries**.
3. Click **Create Text Library**.
4. In the **Create Custom Text Library** dialog box, set the parameters based on the [Parameters for creating a custom text library](#) table. Then, click **OK**.

Create Text Library
✕

\* Name

\* Scene

\* Type  Keyword ⓘ  Similar Text ⓘ

\* Match Mode  Precise ⓘ  Fuzzy ⓘ

\* Category

bizType ⓘ   
[Create BizType](#)

Parameters for creating a custom text library

Parameter	Description
<b>Name</b>	The name of the custom text library. You can set the same name for multiple text libraries. However, we recommend that you set a unique name for each text library.
<b>Scene</b>	The scenario of the text library. Valid values: <ul style="list-style-type: none"> <li>◦ <b>Text Anti-spam</b>: <code>text anti-spam</code> where the value of the scenes parameter contains antispam in API requests</li> <li>◦ <b>Ad</b>: <code>image moderation</code> where the value of the scenes parameter contains ad in API requests</li> </ul>
<b>Type</b>	The text type of the text library. Valid values: <ul style="list-style-type: none"> <li>◦ <b>Keyword</b>: matches the text to be moderated that contains terms. You can detect more risky text by using terms.</li> <li>◦ <b>Similar Text</b>: matches the text to be moderated that is similar to text patterns at a specific probability. You can detect risky text more accurately by using text patterns.</li> </ul> <div style="background-color: #e0f2f7; padding: 5px; margin-top: 10px;"> <p><span>ⓘ</span> <b>Note</b> You can set this parameter to Similar Text only if the Scene parameter is set to Text Anti-spam.</p> </div>

Parameter	Description
<p><b>Match Mode</b></p>	<p>The match mode applied to the custom text library. This parameter is required if the <b>Type</b> parameter is set to <b>Keyword</b>. Valid values:</p> <ul style="list-style-type: none"> <li>◦ <b>Precise</b>: matches the text to be moderated that contains the same terms in the text library.</li> <li>◦ <b>Fuzzy</b>: preprocesses the text to be moderated and terms, and then matches the preprocessed text to be moderated that hits the preprocessed terms. The text to be moderated and terms are preprocessed in the following ways: <ul style="list-style-type: none"> <li>▪ Convert uppercase letters to lowercase letters. For example, if the text to be moderated is "bitCoin", the term "bitcoin" is hit.</li> <li>▪ Convert traditional Chinese characters to simplified Chinese characters.</li> <li>▪ Convert similar words. For example, if the text to be moderated is "②", the term "2" is hit.</li> </ul> </li> </ul> <div style="background-color: #e0f2f1; padding: 5px; margin-top: 10px;"> <p> <b>Note</b> By default, the fuzzy match mode is selected for libraries that consist of text patterns.</p> </div>
<p><b>List Category</b></p>	<p>The category of the moderation result that is returned based on the custom text library.</p> <ul style="list-style-type: none"> <li>◦ Valid values if the <b>Type</b> parameter is set to <b>Keyword</b> parameter: <ul style="list-style-type: none"> <li>▪ <b>Blacklist</b>: If the text to be moderated hits terms in the text library, the machine-assisted moderation result contains the suggestion parameter that has a value of block.</li> <li>▪ <b>Review List</b>: If the text to be moderated hits terms in the text library, the machine-assisted moderation result contains the suggestion parameter that has a value of review.</li> <li>▪ <b>Ignore List</b>: The text excluding that hits terms in the text library is moderated.</li> </ul> </li> <li>◦ Valid values if the <b>Type</b> parameter is set to <b>Similar Text</b>: <ul style="list-style-type: none"> <li>▪ <b>Blacklist</b>: If the text to be moderated hits text patterns in the text library, the machine-assisted moderation result contains the suggestion parameter that has a value of block.</li> <li>▪ <b>Review List</b>: If the text to be moderated hits text patterns in the text library, the machine-assisted moderation result contains the suggestion parameter that has a value of review.</li> <li>▪ <b>Whitelist</b>: If the text to be moderated hits text patterns in the text library, the machine-assisted moderation result contains the suggestion parameter that has a value of pass.</li> </ul> </li> </ul>

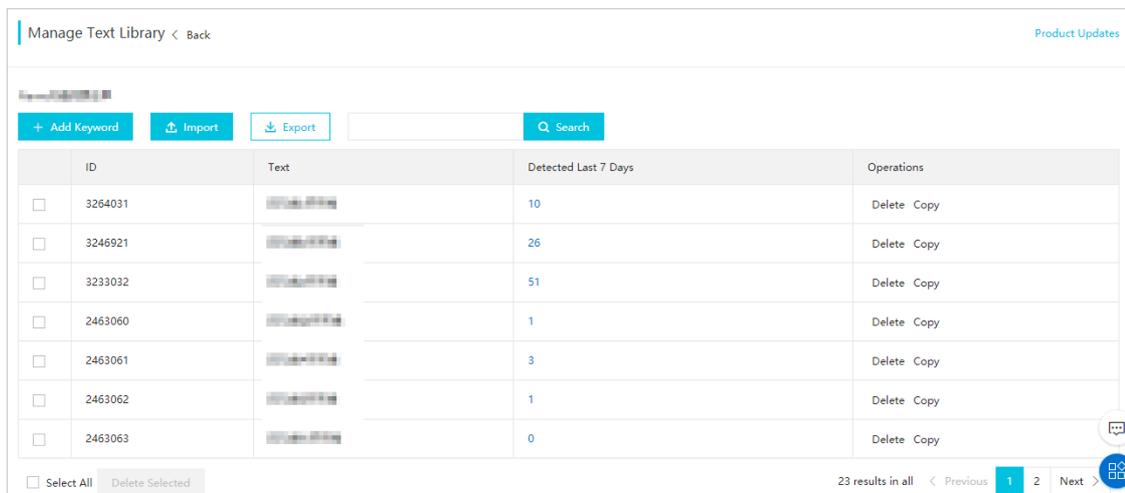
Parameter	Description
<b>bizType</b>	<p>The business scenario to which the custom text library applies. You can specify different text libraries in API requests to meet your business needs. For example, you can use the bizType parameter to specify the text library to be applied in a specific moderation scenario. The bizType parameter takes effect in the following ways:</p> <ul style="list-style-type: none"><li>◦ If the bizType parameter in a moderation request is set to A, the text libraries of which the bizType parameter is set to A are used for moderation. These text libraries can be used only if they are enabled.</li><li>◦ In other cases, all enabled text libraries are used for moderation.</li></ul>

After the text library is created, you can view it in the text library list.

5. Manage terms or text patterns in the text library.

The **Custom Text Library** tab displays all custom text libraries. The libraries marked with **System** and named in the SCENARIO\_FEEDBACK\_WHITE or SCENARIO\_FEEDBACK\_BLACK format are feedback-based text libraries. For example, the **ANTISPAM\_FEEDBACK\_BLACK** library is a blacklist that consists of text patterns added by the system and is used for text anti-spam.

- i. Find the text library that you want to manage and click **Manage** in the Actions column.
- ii. On the **Text Libraries** page, manage terms or text patterns in the library. In this example, you can manage terms.



The **Text Libraries** page displays all text patterns added to the library and displays the number of times each text pattern is hit in the last seven days in the **Detected in Last Seven Days** column, excluding the statistics on the current day.

**Note** You can add and delete terms. The operations take effect in about 15 minutes.

- Click **Add Keyword** or **Import** and add terms as prompted.

- Select one or more terms that you no longer need and click **Batch Delete** to delete the terms. You can also find a specific term and click **Delete** in the Actions column to delete the term.

## Delete, modify, or disable a text library

On the **Custom Text Library** tab, you can click **Delete**, **Edit**, or **Disable** in the **Actions** column to delete, modify, or disable a self-managed text library.

# 1.4. Review machine-assisted moderation results

The Content Moderation console displays the moderation results that are returned after Content Moderation API operations are called. You can use the human review feature to review the machine-assisted moderation results as needed. After you review the machine-assisted moderation result of specific content, the Content Moderation API start to moderate the same content based on the human review results that you set. This topic describes how to review the machine-assisted moderation results.

## Context

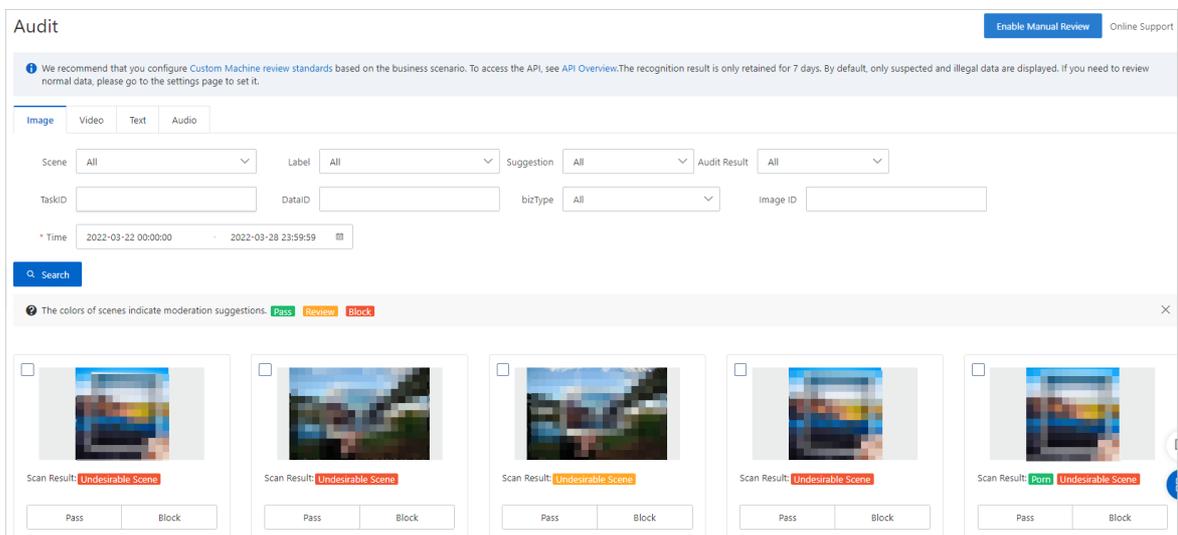
- By default, the Alibaba Cloud Content Moderation console displays only the result data whose machine-assisted moderation result is review or block on the Audit page for human review. If you want to view the result data whose machine-assisted moderation result is pass on the Audit page, you can modify the settings in the console. For more information, see [Customize policies for machine-assisted moderation](#).
- During human review, you can review images, videos, and text. Only the images and text that you review can be automatically added to sample libraries.

- The Alibaba Cloud Content Moderation console retains the machine-assisted moderation results for up to seven days. We recommend that you review the result data at the earliest opportunity.

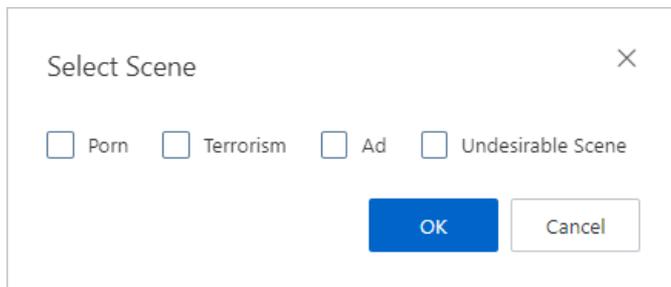
## Procedure

1. Log on to the [Content Moderation console](#).
2. In the left-side navigation pane, choose **Machine audit > Audit**.
3. On the Audit page, click the **Image**, **Video**, or **Text** tab based on the type of data to be reviewed. The corresponding review tab appears.
4. Use the following methods to review the result data:

**Note** This example shows how to review images. The methods for reviewing other types of data are similar.



- You can click **Pass** for the images that have a moderation result of review or block you consider as normal content.
- You can click **Block** for the images that have a moderation result of review or pass but you consider as violations. Then, select **Porn**, **Terrorism**, **Ad**, or **Undesirable Scene** in the dialog box that appears.



- You can select multiple images and review them at a time. To do this, select images as required and click **Batch Mark Pass** or **Batch Mark Block**.
- You can click an image to view its details.



Detail		
URL	https://aligreen-shanghai.oss-cn- [redacted]	Copy
Request Time	2019-12-11 06:03:28 +0800	Copy
TaskID	img7uGkd5yTXya50\$DBzE [redacted]	Copy
DataID	test10e\$Y3jLA1\$6sK7SGc1 [redacted]	Copy
bizType	smoke_ad_program	Copy

The moderation results of reviewed images and similar images are changed based on your human review results. The images with changed moderation results are added to corresponding sample libraries based on the human review results.

## 1.5. View statistics

You can view the statistics about the Content Moderation API in the Alibaba Cloud Content Moderation console.

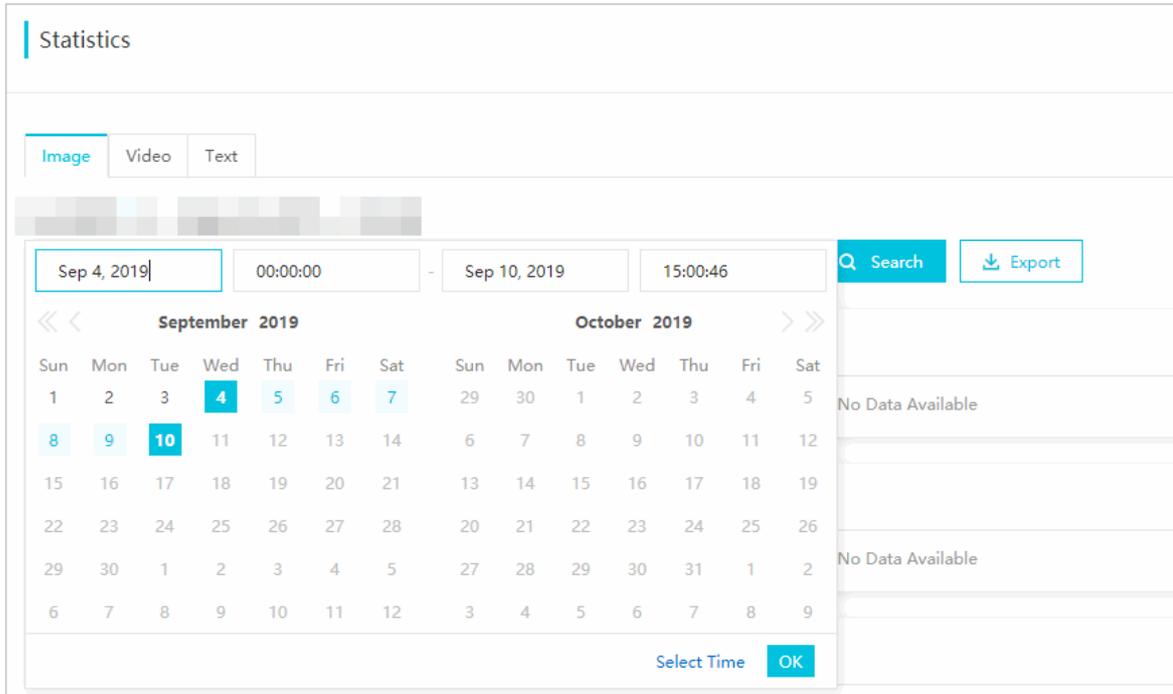
### Context

The Alibaba Cloud Content Moderation console collects statistics on the Content Moderation API. You can query the number of times that the API is called to moderate images, videos, and text in the last year. You can also query the respective numbers of violations, suspected violations, and normal results in different moderation scenarios.

### Procedure

1. Log on to the [Alibaba Cloud Content Moderation console](#).
2. In the left-side navigation pane, choose **Moderation API > Statistics**.
3. On the Statistics page that appears, click the **Image**, **Video**, or **Text** tab based on the type of the data to be queried.
4. On the **Statistics** page, specify the time range that you want to query and click **Search**.

You can query the data in the last year. You can query data for a maximum duration of a month.

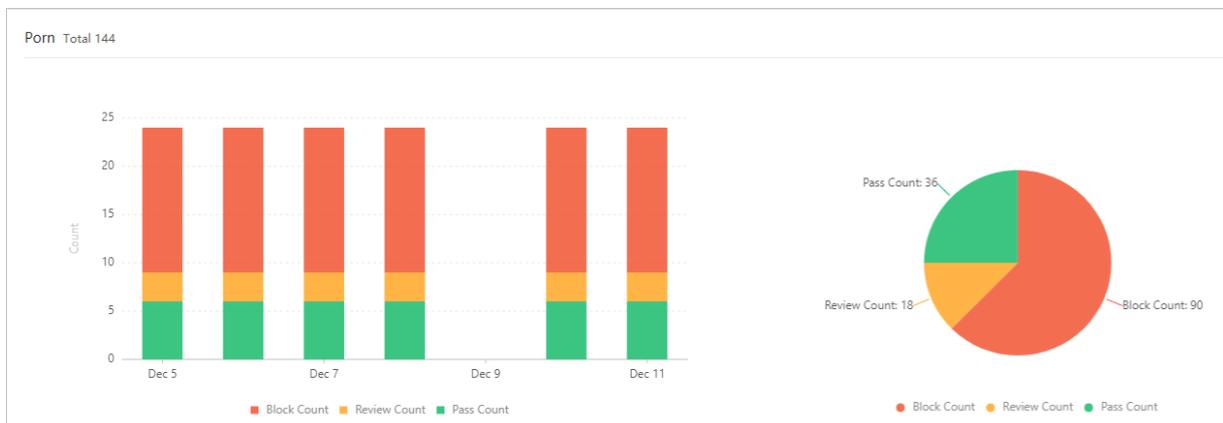


- 5. View the statistics in different moderation scenarios and export the reports as needed.
  - o [Reports in the console](#)
  - o [Exported reports](#)

## Reports in the console

For example, the **Porn** report shows the daily number of times that the Content Moderation API is called to moderate images for pornographic content, that is, the value of the scenes parameter contains porn in API requests. The report also shows the respective numbers of violations, suspected violations, and normal results. The legends are described as follows:

- **Block Count** : the number of moderation requests for which the value of the suggestion parameter is block in the returned moderation results.
- **Review Count** : the number of moderation requests for which the value of the suggestion parameter is review in the returned moderation results.
- **Pass Count** : the number of moderation requests for which the value of the suggestion parameter is pass in the returned moderation results.



The unit of the call volume varies with the moderation object :

- Image moderation: the number of images.
- Video moderation: the number of captured frames for a video or the duration of a video in units of minutes.
- Text moderation: the number of text entries.

## Exported reports

You can export the reports in Excel format. The time range of each exported report is consistent with the time range that you set to query data. Each exported report in Excel format contains the statistics about the Content Moderation API that is called in a specific moderation scenario based on the value of the scenes parameter in API requests. The statistics are collected on a daily basis.

The following table describes the headers in the exported reports.

Header	Description	Unit
day	The date on which the API is called.	None
totalImageCount	The total number of images that are moderated.	Piece
blockImageCount	The number of images that contain violations.	Piece
reviewImageCount	The number of images that contain suspected violations.	Piece
passImageCount	The number of images that are normal.	Piece
totalVideoCount	The total number of videos that are moderated.	Piece
blockVideoCount	The number of videos that contain violations.	Piece
reviewVideoCount	The number of videos that contain suspected violations.	Piece
passVideoCount	The number of videos that are normal.	Piece
innerFrameCount	The number of video frames captured by Content Moderation.	Piece
outerFrameCount	The number of video frames submitted by users.	Piece
totalTextCount	The total number of text entries that are moderated.	Piece
blockTextCount	The number of text entries that contain violations.	Piece
reviewTextCount	The number of text entries that contain suspected violations.	Piece

Header	Description	Unit
passTextCount	The number of text entries that are normal.	Piece

## 1.6. Give feedback on samples

If the machine-assisted moderation results returned by Content Moderation do not meet your expectations, you can call feedback operations to give feedback on samples.

If you use the Content Moderation API on your own platform, you can also call feedback operations to give feedback on samples whose machine-assisted moderation results do not meet your expectations. Alibaba Cloud can add the feedback data to model training and correct the moderation results in later versions. The trained model can be more adaptive to your business scenarios.

Currently, Content Moderation allows you to give feedback on [image samples](#), [video samples](#), and [text samples](#). For more information, see the API reference.

### Automatically add samples to a custom image library

Model training requires sufficient samples and takes some time. You can enable Content Moderation to automatically add samples to a custom image library and correct the moderation results in real time.

For the feedback on image samples, you can manage feedback-based image libraries in the Alibaba Cloud Content Moderation console. The operations to manage feedback-based image libraries are the same as those to manage user-created image libraries, except that you cannot create or delete a feedback-based image library. For more information, see [Manage custom image libraries](#).

To automatically add samples to a custom image library, you can set the label parameter when you call feedback operations:

- For normal images, set this parameter to *normal* to add these images to the whitelist.
- For risky images, set this parameter to other values to add these images to the blacklist. We recommend that you use values that indicate risks such as *porn*, *ad*, and *terrorism*.

## 1.7. Authorize a role to access ApsaraVideo for Media Processing

Content Moderation calls ApsaraVideo for Media Processing to capture frames for the videos that are uploaded to Object Storage Service (OSS). You must authorize a role to access ApsaraVideo for Media Processing. Then, Content Moderation can use the role to submit a video frame capture task. You must use the role management feature of Alibaba Cloud Resource Access Management (RAM). This topic describes how to perform the required steps.

### Context

- If you submit an asynchronous video moderation task and specify the video URL in the `oss://xxxx` format, Content Moderation can automatically capture frames for the video that is uploaded to OSS. To ensure automatic frame capture, you must authorize a role to access ApsaraVideo for Media Processing.
- The OSS service is available in the China (Hangzhou), China (Shanghai), China (Beijing), and China (Shenzhen) regions.

### Procedure

1. Create a RAM role and grant permissions.

Create a RAM role that can access ApsaraVideo for Media Processing for your Alibaba Cloud account. Authorize your Alibaba Cloud account to assume the role.

- i. Log on to the [RAM console](#).
- ii. Go to the **RAM Roles** page. Then, click **Create RAM Role**.
- iii. Select **Alibaba Cloud Service** as the trusted entity type and click **Next**.
- iv. Set the **RAM Role Name** parameter and select **Media Transcoding Service** from the Select Trusted Service drop-down list.
- v. Click **OK**.

 **Note** This operation may require mobile phone-based authentication.

- vi. After the role is created, click **Add Permissions to RAM Role**.
  - vii. In the **Add Permissions** panel, select the **AliyunOSSReadOnlyAccess** policy and click **OK**. This operation grants the RAM role only read permissions on OSS data that belongs to your Alibaba Cloud account.
2. Modify the trust policy of the RAM role.
    - i. On the **RAM Roles** page, click the name of the created role to go to the role details page.
    - ii. Click the **Trust Policy Management** tab and then click **Edit Trust Policy**.
    - iii. In the **Edit Trust Policy** panel, replace the content of `"Service"` with `"1184847062244573@mts.aliyuncs.com"`. Click **OK**.  
This operation specifies that the Alibaba Cloud account whose UID is 1184847062244573 assumes the created RAM role to call ApsaraVideo for Media Processing. Content Moderation is also used within this Alibaba Cloud account.

## What's next

After the role is authorized, you must use a URL of the specified format when you submit an asynchronous moderation task for a video to be moderated in OSS. To use the URL of the specified format, perform the following steps:

1. On the role details page, view and copy the Alibaba Cloud resource name (ARN).
2. Generate the video URL for a video to be moderated in OSS in the following format: `oss://arn@bucket.region/object`.

Assume that you want to moderate the `video/bar.mp4` object in the `f**` OSS bucket that resides in the China (Shenzhen) region. The generated URL of the video is `oss://acs:ram:xxxxxxxxxxxxxxxx:role/mts-to-a@f**,cn-shenzhen/video/bar.mp4`. `xxxxxxxxxxxxxxxx` is the 16-digit ID of your Alibaba Cloud account.

3. When you submit an asynchronous video moderation task, use the generated video URL.

### Note

- The method of submitting a video URL to specify the video to be moderated is available only for asynchronous video moderation tasks. For more information, see [Asynchronous moderation](#).
- The supported regions are China (Hangzhou), China (Shanghai), China (Beijing), and China (Shenzhen).

## 2. RAM user permissions

### 2.1. Grant read-only permissions of Content Moderation to a RAM user

You can use Alibaba Cloud Content Moderation as a RAM user. By default, Alibaba Cloud Resource Access Management (RAM) provides only the AliyunYundunGreenWebFullAccess system policy for Content Moderation. You can attach this system policy to grant a RAM user full management permissions of Content Moderation. If you need to grant a RAM user only read-only permissions of Content Moderation, you can create a custom policy to grant permissions. Then, this RAM user cannot perform write operations, for example, configure the scan settings for Object Storage Service (OSS) violation detection and manage the scanning results of OSS violation detection in the Alibaba Cloud Content Moderation console.

#### Prerequisites

A RAM user is created. For more information, see [Create a RAM user](#).

#### Procedure

1. Log on to the [Alibaba Cloud RAM console](#).
2. Create a custom policy for read-only permissions of Content Moderation.
  - i. In the left-side navigation pane, choose **Permissions > Policies**.
  - ii. Click **Create Policy**.
  - iii. On the **Create Custom Policy** page, enter a policy name in the **Policy Name** field, select **Script for Configuration Mode**, and then enter the following script:

```
{
  "Version": "1",
  "Statement": [{
    "Action": [
      "yundun-greenweb:List*",
      "yundun-greenweb:Get*",
      "yundun-greenweb:Describe*",
      "yundun-greenweb:Query*"
    ],
    "Resource": "*",
    "Effect": "Allow"
  }]
}
```

RAM / Policies / Create Custom Policy

## ← Create Custom Policy

\* Policy Name

Note

Configuration Mode

Visualized

Script

Policy Document

```
2  "Version": "1",
3  "Statement": [{
4    "Action": [
5      "yundun-greenweb:List*",
6      "yundun-greenweb:Get*",
7      "yundun-greenweb:Describe*",
8      "yundun-greenweb:Query*"
9    ],
10   "Resource": "*",
11   "Effect": "Allow"
```

OK Back

iv. Click **OK**.

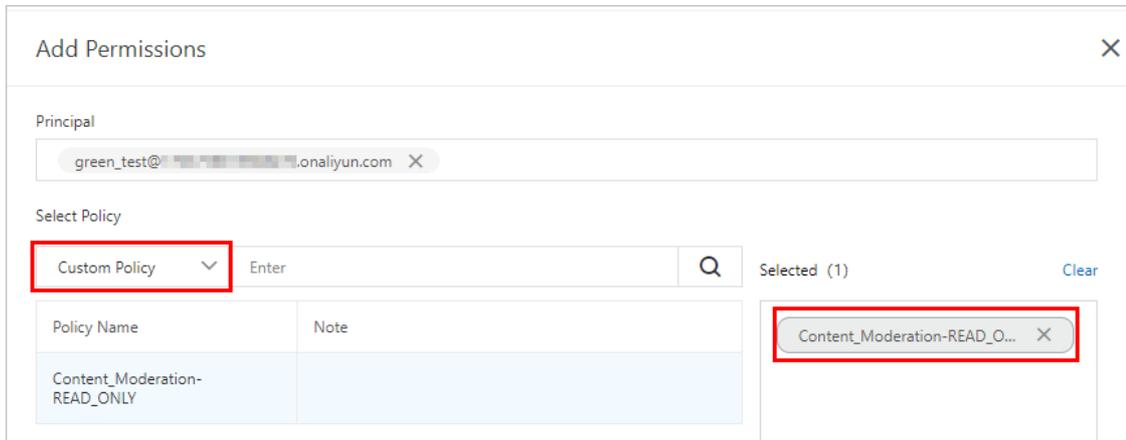
The custom policy is created.

3. Grant a RAM user read-only permissions of Content Moderation.

i. In the left-side navigation pane, choose **Identities > Users**.

ii. On the **Users** page, find the user who you want to grant permissions and click **Add Permissions** in the Actions column.

- iii. In the **Add Permissions** pane, click **Custom Policy** in the **Select Policy** section, search for and select the custom policy that you create in Step 2, and then click **OK**.

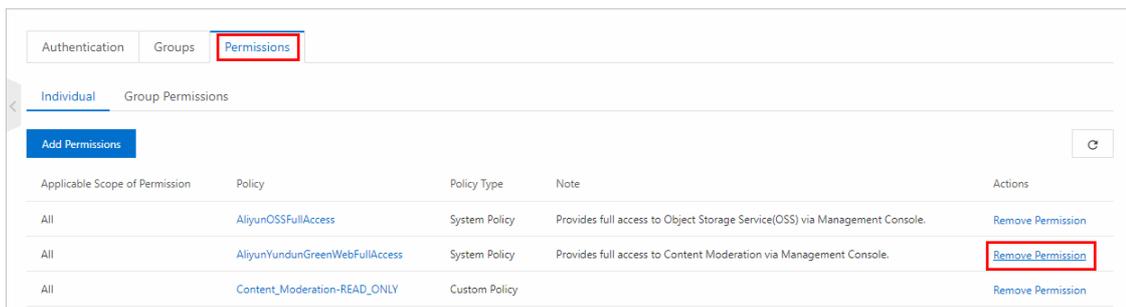


- iv. Confirm the authorization result and click **Complete**.

Read-only permissions of Content Moderation are granted to the RAM user.

4. (Optional) If the **AliyunYundunGreenWebFullAccess** system policy is attached to the RAM user who you want to grant only read-only permissions, remove this system policy to revoke full management permissions of Content Moderation from the RAM user.
  - i. On the **Users** page, click the name of the RAM user.
  - ii. On the user details page, click **Permissions**.
  - iii. On the **Individual** tab, find the **AliyunYundunGreenWebFullAccess** system policy and click **Remove Permission** in the **Actions** column.

**Note** If the RAM user inherits the **AliyunYundunGreenWebFullAccess** permission from a user group, you must remove the system policy from the group or remove the RAM user from the group. You can check whether the RAM user inherits the **AliyunYundunGreenWebFullAccess** permission on the **Group Permissions** tab. For more information, see [Remove permissions from a RAM user group](#) and [Remove a RAM user from a RAM user group](#).



- iv. In the message that appears, click **OK**.

The system policy is removed.

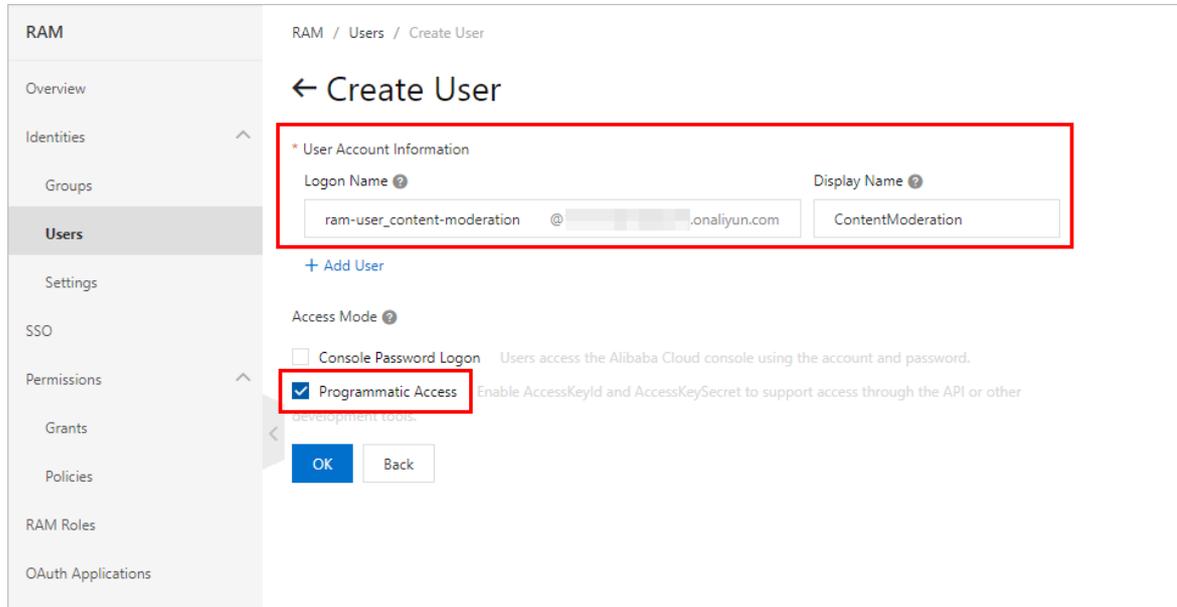
## 2.2. Call the Content Moderation API as a RAM user

You can call the Content Moderation API as a RAM user. Before calling the Content Moderation API, make sure that you have created a RAM user and granted relevant permissions to the RAM user.

## Procedure

1. Log on to the [RAM console](#), create a RAM user, and then generate the AccessKey ID and AccessKey secret. Properly keep the generated AccessKey ID and AccessKey secret for calling the SDKs later.

For more information about how to create a RAM user, see [Create a RAM user](#).



2. Grant relevant permissions to the RAM user. Only an authorized RAM user can call the Content Moderation API. Grant the following system policy permission to the RAM user you created: `AliyunYundunGreenWebFullAccess`.

For more information about how to grant permissions to a RAM user, see [Grant permissions to a RAM user](#).

