Alibaba Cloud

内容安全 FAQ

Document Version: 20220602

C-J 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.

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

Table of Contents

1.Use the service	05
1.1. FAQ about Content Moderation features	05
1.2. FAQ about the Content Moderation API	80
1.3. FAQ about OCR	18
2.FAQ about billing	19
3.FAQ about error messages	20
4.FAQ about moderation effects	28

1.Use the service

1.1. FAQ about Content Moderation features

This topic provides answers to the frequently asked questions (FAQ) about Content Moderation features.

- What are the features of Content Moderation?
- What is a custom text library of Content Moderation?
- Can I customize configurations for image moderation in Content Moderation?
- How do I use the human review feature in Content Moderation?
- Why does the custom text library that I configured in Content Moderation not take effect?
- Can Content Moderation detect spelling or grammatical errors?
- What is the scope of terrorist content detection in Content Moderation?
- What does the rate parameter indicate in image moderation?
- Can Lexport a custom text library from Content Moderation?
- When will the human review results take effect in the Content Moderation background?
- How long does a custom image or text library take effect in Content Moderation?
- Why am I unable to receive callback notifications after a callback URL is specified in Content Moderation?
- Why am I unable to moderate a long image after I add long image samples to a custom image library?
- Why does a custom policy for machine-assisted moderation configured in the Content Moderation console not take effect when I call a Content Moderation operation?

What are the features of Content Moderation?

Content Moderation provides the Content Moderation API, Object Storage Service (OSS) violation detection, and site detection features that apply to various scenarios.

For more information, see Functions and features.

What is a custom text library of Content Moderation?

Content Moderation supports custom text libraries of different types. You can use custom text libraries to ensure that moderation results meet specific business needs.

For more information, see Manage custom text libraries.

Can I customize configurations for image moderation in Content Moderation?

Yes, Content Moderation supports custom image libraries. You can use custom image libraries to ensure that moderation results meet specific business needs.

For more information, see Manage custom image libraries.

How do I use the human review feature in Content Moderation?

The Content Moderation console displays the moderation results that are returned after Content Moderation operations are called. You can use the human review feature to review the machine-assisted moderation results again as needed.

For more information, see Review machine-assisted moderation results.

Why does the custom text library that I configured in Content Moderation not take effect?

Custom text libraries in Content Moderation include term libraries and text pattern libraries. Matching methods include exact match and fuzzy match. Check whether your custom text library meets the requirements for custom text libraries. For example, make sure that each text pattern in the custom text library contains at least 10 characters. If you use the custom text library when you call Content Moderation operations, make sure that the custom text library is applied to the corresponding business scenario. Otherwise, the custom text library does not take effect.

For more information about how to use custom text libraries, see Manage custom text libraries.

Can Content Moderation detect spelling or grammatical errors?

Content Moderation cannot detect spelling or grammatical errors.

What is the scope of terrorist content detection in Content Moderation?

Terrorist content detection allows you to moderate objects such as images for terrorist content, including bloody content, explosion and smoke, special costumes, logos, weapons, political content, violence, crowds, parades, car accidents, flags, and landmarks. In actual content moderation, you can define the moderation scope of your business by customizing policies for machine-assisted moderation.

For more information, see Customize policies for machine-assisted moderation.

What does the rate parameter indicate in image moderation?

We recommend that you determine whether an image contains violations based on the values of the suggestion and label parameters instead of the value of the rate parameter. The rate parameter indicates only the confidence level that the Content Moderation model calculates for an image. It does not reflect the risk level of an image.

For more information about the parameters, see Synchronous moderation.

Can I export a custom text library from Content Moderation?

Yes, you can export multiple terms from a custom text library at a time in the Content Moderation console.

For more information, see Manage custom text libraries.

When will the human review results take effect in the Content Moderation background?

You can use the human review feature to review machine-assisted moderation results in the Content Moderation console. The human review results take effect in real time. The review results of images and text are automatically added to sample libraries. The newly added samples in the sample libraries take about 15 minutes to be effective.

For more information, see Review machine-assisted moderation results.

How long does a custom image or text library take effect in Content Moderation?

Content Moderation supports custom image libraries. You can use custom image libraries to manage the images that you want to block or skip. You can add and remove image samples in custom image libraries. All these operations take effect about 15 minutes later after they are performed. Content Moderation also supports custom text libraries. You can use custom text libraries to manage the text that you want to block or skip. You can add and remove text patterns and terms in custom text libraries. All these operations take effect about 15 minutes later after they are performed.

For more information, see Manage custom text libraries.

Why am I unable to receive callback notifications after a callback URL is specified in Content Moderation?

Both the Content Moderation API and the OSS violation detection feature support callback notifications. When you create a notification plan in the Content Moderation console, you must specify the notification type such as machine-assisted moderation results, human review results, or manual review results. Then, you must associate the created notification plan with the corresponding business scenario before the notification plan takes effect.

If you still cannot receive a callback notification from Content Moderation after you create a valid notification plan, we recommend that you check whether your server can respond to the POST requests sent to the specified callback URL. This ensures that no 403 or 502 error occurs when the callback URL is requested. You can also set the callback parameter in an API request for content moderation to specify a callback URL. After a Content Moderation operation is called, Content Moderation sends a callback notification to the specified callback URL.

For more information, see Enable callback notifications.

Why am I unable to moderate a long image after I add long image samples to a custom image library?

Content Moderation can moderate images whose height does not exceed 400 pixels or whose aspect ratio does not exceed 2.5. For a long image whose height or aspect ratio exceeds the upper limit, Content Moderation crops the long image and then moderates it. As a result, the cropped long image does not hit the long image samples added to a custom image library or feedback-based image library. Therefore, the long image cannot be moderated.

For more information about the **interval** and **maxframes** parameters for image moderation, see Synchronous moderation.

Why does a custom policy for machine-assisted moderation configured in the Content Moderation console not take effect when I call a Content Moderation operation?

After you customize a policy for machine-assisted moderation or add a sample to a custom sample library in the Content Moderation console, the custom policy or sample takes 15 minutes to be effective. We recommend that you try again later. In addition, after you customize a policy for machine-assisted moderation based on a business scenario, you must specify the business scenario in an API request for content moderation. Then, the corresponding moderation policy takes effect.

For more information, see Customize policies for machine-assisted moderation.

1.2. FAQ about the Content Moderation API

This topic provides answers to the frequently asked questions (FAQ) about the Content Moderation API.

- How can I query the statistics on the Content Moderation API?
- Which operation can I call to moderate text in Content Moderation?
- How do I give feedback on the errors in text moderation results in Content Moderation?
- How can I view the descriptions of the parameters that are returned after a Content Moderation operation is called?
- Can I include the signature information of an API request in the request body?
- Can illegal links in the text be detected during text moderation?
- How long does the moderation duration of a Content Moderation operation last?
- Can I use Composer to download Content Moderation SDK for PHP?
- Can I call the /green/text/scan operation to moderate English text in Content Moderation?
- Can I call a video moderation operation to moderate a video whose size is larger than 2 GB in Content Moderation?
- What permissions does a RAM user require to call a Content Moderation operation? How are the permissions granted?
- What is the size limit of an image to be moderated in Content Moderation?
- Can I submit the internal URLs of objects for moderation?
- How do I moderate images that exceed the size limit?
- Can I create HTTP requests to call Content Moderation operations?
- Can I use the AccessKey pair of a RAM user to call Content Moderation operations?
- Does the Content Moderation API provide call examples?
- Can I use an SDK for .NET to call Content Moderation operations?
- Can Content Moderation moderate images in the GIF format?
- Can I extend the maximum download duration from 3 seconds to a longer period of time when I call a Content Moderation operation?
- How many images can be moderated at most each time the /green/image/scan operation is called in Content Moderation?
- Can I call a single Content Moderation operation to simultaneously moderate content in multiple scenarios, such as pornography detection and terrorist content detection?
- What domain names and ports are available for calling Content Moderation operations?
- How can I call a video moderation operation in Content Moderation to moderate a video in ApsaraVideo VOD?
- Can Content Moderation moderate M3U8 video files?
- How long does an asynchronous task for video moderation take?
- Can I call Content Moderation operations in a region of the United States to moderate videos?
- What are the differences between the /green/video/syncscan and /green/video/asyncscan operations in Content Moderation?
- Can I set the callback parameter in an API request for asynchronous image moderation by using Content Moderation SDK for Java?
- What does the BizType parameter specify in Content Moderation?

内容安全

- What is the purpose of associating a text library with multiple business scenarios in Content Moderation?
- Why is the value of the checksum parameter in the callback notification different from the calculated value after I call the /green/video/asyncscan operation?
- Why is no data ID returned after I call the /green/text/scan operation in Content Moderation?
- Why are different labels returned for the same image in single-scenario moderation and multiscenario moderation?
- Why is the context parameter not returned in text moderation results?
- Why is the filteredContent parameter returned but the context parameter not returned in text moderation results?
- Why do text moderation results contain no emojis?
- Do the accuracy rate and detection rate of synchronous image moderation differ from those of asynchronous image moderation in Content Moderation?
- Why am I unable to download the Client Uploader utility class that is used to moderate local files and binary files for Content Moderation SDK for Java?
- Why does the aliyunsdkcore library fail to be installed for Content Moderation SDK for Python 3.5.4 and 3.8.8?
- How do I install the aliyunsdkgreenextension utility class of Content Moderation SDK for Python?
- Can a custom term library in Content Moderation contain terms in languages other than English?
- How long is the validity period of an OSS URL for an image or a video that is uploaded from a local machine to OSS for content moderation?
- How am I charged for moderating a video live stream in Content Moderation?
- How are the moderation results of a live stream returned to the callback URL in Content Moderation?
- Why does my callback URL still receive data after Content Moderation stops moderating a live stream?
- When will the status code 200 be returned during live stream moderation?

How can I query the statistics on the Content Moderation API?

The Content Moderation console collects statistics on the Content Moderation API. You can query the number of times that Content Moderation operations were called to moderate images, videos, and text over the last year.

For more information, see View statistics.

Which operation can I call to moderate text in Content Moderation?

You can call the /green/text/scan operation to detect whether the specified Chinese or English text contains violations.

For more information about the /green/text/scan operation, see Text Synchronous moderation.

How do I give feedback on the errors in text moderation results in Content Moderation?

If you find that the text moderation results contain errors, you can call the /green/text/feedback operation to give feedback on the errors.

For more information, see Give feedback on moderation results.

How can I view the descriptions of the parameters that are returned after a Content Moderation operation is called?

You can view the common response parameters and common HTTP status codes that are returned after a Content Moderation operation is called.

For more information, see Common parameters.

Can I include the signature information of an API request in the request body?

No, the signature information must be included in the request header. You must set the signature parameter in the Authorization header of an HTTP request to specify the signature information for verification. We recommend that you use Content Moderation SDKs. Content Moderation supports SDKs for various programming languages such as Java, Python, and PHP.

For more information, see Common parameters and SDK overview.

Can illegal links in the text be detected during text moderation?

No, Content Moderation can detect violations, but not illegal links, in the text during text moderation.

How long does the moderation duration of a Content Moderation operation last?

The moderation duration of different Content Moderation operations varies.

- Images: An image takes about 300 milliseconds to moderate, excluding the download duration.
- Videos:
 - Video files: A video file can be moderated at a speed about one to six times the playback speed, excluding the download duration. For example, the moderation speed is six times the playback speed. In this case, a 6-minute video takes 1 minute to moderate.
 - Video streams: The moderation duration of video streams varies based on the frame rate. In general, moderation results are returned within a second after a frame is captured.
- Text: In general, moderation results are returned within 50 milliseconds.

Can I use Composer to download Content Moderation SDK for PHP?

Yes, you can use Composer to download Content Moderation SDK for PHP, provided that your Composer supports PHP 5.3 or later. No tutorial is provided for installing Composer.

For more information about how to install Content Moderation SDK for PHP, see Installation.

Can I call the /green/text/scan operation to moderate English text in Content Moderation?

Yes, you can call the /green/text/scan operation to moderate English text.

For more information, see Text Synchronous moderation.

Can I call a video moderation operation to moderate a video whose size is larger than 2 GB in Content Moderation?

By default, the size of a single video to be moderated in Content Moderation cannot exceed 200 MB. You can contact the technical support to raise the size limit up to 2 GB as needed. If you want to moderate a video whose size is larger than 2 GB, we recommend that you segment the video into multiple video segments and then moderate them. By default, 200 frames are captured from a video. If you want to capture more frames from a large video for moderation, you must set the maxFrames parameter. By default, a maximum of 3,600 frames can be captured.

For more information, see Asynchronous moderation.

What permissions does a RAM user require to call a Content Moderation operation? How are the permissions granted?

You can use the AccessKey ID and AccessKey secret of a RAM user to call a Content Moderation operation. Before you call a Content Moderation operation as a RAM user, permissions must be granted to the RAM user.

For more information, see Call the Content Moderation API as a RAM user.

What is the size limit of an image to be moderated in Content Moderation?

Content Moderation can moderate an image of which the size does not exceed 20 MB, the height or width does not exceed 30,000 pixels, and the total resolution does not exceed 0.25 billion pixels.

For more information, see Synchronous moderation.

What is a concurrency limit for calling Content Moderation operations?

A concurrency limit specifies the total number of images, videos, or text entries that can be moderated at the same time. This limit applies to both the pay-as-you-go billing method and subscription plans.

The following table describes the concurrency limits for calling Content Moderation operations to moderate different types of objects.

Moderation object	Default concurrency limit	Unit	Description
lmage	50	N/A	The maximum number of images that can be moderated per second.
Video	20	N/A	The maximum number of videos that can be moderated at the same time. Files and streams are not differentiated.
Text	100	Entry	The maximum number of text entries that can be moderated per second. Each text entry contains less than 200 characters.

? Note

- Images, videos, or text entries moderated within the default concurrency limit are free of charge. To adjust the default concurrency limit, . If you raise the default concurrency limit, you will have extra charges.
- By default, if an object is simultaneously moderated in multiple scenarios, only one object is counted. For example, if you send an API request to moderate an image for pornography and terrorist content at the same time, only one image is counted.

Can I submit the internal URLs of objects for moderation?

No, only public URLs are supported for moderation. If you worry about the risks of data leaks, we recommend that you set a short validity period for the public URLs, such as 10 minutes.

How do I moderate images that exceed the size limit?

We recommend that you compress the images before you submit them for moderation. When the resolution of an image is greater than 256 × 256 pixels, the resolution has little impact on the moderation results.

Can I create HTTP requests to call Content Moderation operations?

Yes, you can create HTTP requests to call Content Moderation operations. However, you must sign each request, which is a tedious operation. We recommend that you use the SDKs provided on the Alibaba Cloud official website to call Content Moderation operations.

Can I use the AccessKey pair of a RAM user to call Content Moderation operations?

- You can use the AccessKey pair of a RAM user to call Content Moderation operations whose API version is V20160621 or later. For more information about the required dependencies, see SDK overview.
- You cannot use the AccessKey pair of a RAM user to call Content Moderation operations whose API version is earlier than V20160621. Otherwise, an **AccessDenied** error is returned.

Does the Content Moderation API provide call examples?

Yes, the Content Moderation API provides call examples. You can download the examples in SDK Reference. For more information, see SDK overview.

? Note The version date of a call example changes as the API is updated. We recommend that you check the official documentation on a regular basis.

Can I use an SDK for .NET to call Content Moderation operations?

No, Content Moderation does not provide an SDK for .NET. We recommend that you use Content Moderation SDKs for other programming languages. Alternatively, you can create HTTP requests to call Content Moderation operations.

For more information, see SDK overview and Request structure.

Can Content Moderation moderate images in the GIF format?

Yes, Content Moderation can moderate images in the PNG, JPG, JPEG, BMP, GIF, or WEBP format.

For more information, see Synchronous moderation.

Can I extend the maximum download duration from 3 seconds to a longer period of time when I call a Content Moderation operation?

No, the maximum download duration cannot be extended. If download errors frequently occur when you call a Content Moderation operation to moderate an image, check whether the image URL is accessible or whether the image can be downloaded within 3 seconds. You can select a region nearest to the region where your server resides. This saves time for calling operations.

For more information, see Endpoint.

How many images can be moderated at most each time the /green/image/scan operation is called in Content Moderation?

A maximum of 100 images can be moderated each time. In other words, a maximum of 100 images can be submitted for moderation each time. To moderate 100 images each time, you must raise the default concurrency limit for images. By default, a maximum of 50 images, 100 text entries, or 20 videos can be moderated each time.

For more information, see Synchronous moderation and Pricing.

Can I call a single Content Moderation operation to simultaneously moderate content in multiple scenarios, such as pornography detection and terrorist content detection?

Yes, you can call a single Content Moderation operation to simultaneously moderate content in multiple scenarios. To do this, set the scenes parameter in an API request to specify multiple scenarios. For example, you can set the scenes parameter to **["porn","terrorism"]** to intelligently detect pornography and terrorist content in images. If your moderation involves multiple scenarios at a time, you are charged for all the scenarios. The cost of each scenario equals the number of images that are moderated in the scenario multiplied by the unit price of the scenario.

For more information, see Synchronous moderation and Pricing.

What domain names and ports are available for calling Content Moderation operations?

If you need to configure a network security policy, we recommend that you enable access from ***.aliyuncs.com** and enable ports 80 and 443.

How can I call a video moderation operation in Content Moderation to moderate a video in ApsaraVideo VOD?

To call a video moderation operation in Content Moderation to moderate a video in ApsaraVideo VOD, you cannot directly submit the ID of the video. You must submit a sequence of frames that are captured from the video or a URL that can be used to access the video.

For more information, see Asynchronous moderation.

Can Content Moderation moderate M3U8 video files?

No, Content Moderation cannot moderate **M3U8** video files. Content Moderation can moderate video files in the following formats: *AVI, FLV, MP4, MPG, ASF, WMV, MOV, WMA, RMVB, RM, FLASH,* and *TS*.

For more information, see Asynchronous moderation.

How long does an asynchronous task for video moderation take?

The duration of an asynchronous task for video moderation depends on the download duration of the object to be moderated. In addition, the duration that is required to moderate different types of objects varies.

- Video files: A video file can be moderated at a speed about one to six times the playback speed, excluding the download duration. For example, the moderation speed is six times the playback speed. In this case, a 6-minute video takes 1 minute to moderate.
- Video streams: The moderation duration of video streams varies based on the frame rate. In general, moderation results are returned within a second after a frame is captured.

Can I call Content Moderation operations in a region of the United States to moderate videos?

No, Content Moderation operations cannot be called in a region of the United States. The following regions are supported:

- China (Shanghai): cn-shanghai
- China (Beijing): cn-beijing
- China (Shenzhen): cn-shenzhen
- Singapore (Singapore): ap-southeast-1

If you run your business outside China, we recommend that you use the endpoint of the Singapore (Singapore) region to call Content Moderation operations. For more information, see Endpoint.

What are the differences between the /green/video/syncscan and /green/video/asyncscan operations in Content Moderation?

To call the /green/video/syncscan operation, you must submit a sequence of frames that are captured from the video to be moderated. If you want to submit a video URL to specify the video to be moderated, we recommend that you call the /green/video/asyncscan operation.

The /green/video/asyncscan operation can be called to moderate video files and video streams. To moderate a video file, you can submit a sequence of frames that are captured from the video file or a URL that specifies the video file. However, you cannot obtain the moderation results of an asynchronous moderation task in real time. To obtain the moderation results, you can set the callback parameter in the API request. Alternatively, you can call the /green/video/results operation to poll the moderation results.

For more information, see Synchronous moderation and Asynchronous moderation.

Can I set the callback parameter in an API request for asynchronous image moderation by using Content Moderation SDK for Java?

Yes, you can set the callback parameter in an API request for asynchronous image moderation.

For more information, see Image moderation.

What does the BizType parameter specify in Content Moderation?

The BizType parameter specifies a business scenario. Each business scenario corresponds to a moderation policy. Before you use the Content Moderation API, we recommend that you create business scenarios based on your business requirements. After you customize a moderation policy 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.

For more information, see Customize policies for machine-assisted moderation.

What is the purpose of associating a text library with multiple business scenarios in Content Moderation?

When you create a custom text or image library, we recommend that you associate the custom text or image library with the business scenario to which the library applies. For example, your text library is associated with Business Scenario A, and you have specified Business Scenario A in an API request for text moderation. In this case, the text library that is associated with Business Scenario A is used for text moderation. Otherwise, all enabled text libraries are used for text moderation.

For more information, see Manage custom text libraries.

Why is the value of the checksum parameter in the callback notification different from the calculated value after I call the /green/video/asyncscan operation?

The value of the checksum parameter is a string in the <UID> + <Seed> + <Content> format. It is generated by using the Secure Hash Algorithm 256 (SHA-256) algorithm. UID indicates the ID of your Alibaba Cloud account. You can query the ID in the Alibaba Cloud Management Console. To prevent data in the callback notification from being tampered with, you can use the SHA-256 algorithm to generate a string when your server receives the callback notification. Then, you can verify the string against the received checksum parameter.

For more information, see Enable callback notifications.

Why is no data ID returned after I call the /green/text/scan operation in Content Moderation?

If you have specified a data ID in the API request for the /green/text/scan operation, the data ID is returned after you call the operation.

For more information, see Text Synchronous moderation.

Why are different labels returned for the same image in singlescenario moderation and multi-scenario moderation?

In general, this is because the configurations in multiple scenarios are different from those in a single scenario. Therefore, the configurations conflict when the same image is moderated across scenarios. We recommend that you contact algorithm engineers to check whether the scenario configurations are different. Alternatively, you can moderate the same image separately in different scenarios.

For more information, see Synchronous moderation and Image moderation.

Why is the context parameter not returned in text moderation results?

The context parameter indicates the risky terms that the moderated text hits. The hit terms are returned. If the moderated text hits other policies such as algorithm models or text patterns, this parameter is not returned.

For more information, see Text Synchronous moderation.

Why is the filteredContent parameter returned but the context parameter not returned in text moderation results?

The **filteredContent** parameter indicates the text that is returned after hit terms in the moderated text are redacted with asterisks (*). If the moderated text hits specific terms or text patterns in your custom text library, this parameter is returned. The **context** parameter indicates the risky terms that the moderated text hits. The hit terms are returned. If the moderated text hits other policies such as algorithm models or text patterns, this parameter is not returned.

For more information, see Text Synchronous moderation.

Why do text moderation results contain no emojis?

Content Moderation cannot recognize emojis in text. Emoji characters are filtered out in the returned moderation results.

Do the accuracy rate and detection rate of synchronous image moderation differ from those of asynchronous image moderation in Content Moderation?

No, synchronous image moderation and asynchronous image moderation have the same moderation effects. The only difference is that they are implemented by different operations.

Why am I unable to download the ClientUploader utility class that is used to moderate local files and binary files for Content Moderation SDK for Java?

You must download the Client Uploader utility class for Content Moderation SDK for Java and import it to your project.

For more information about the download URL and procedure, see Installation.

Why does the aliyunsdkcore library fail to be installed for Content Moderation SDK for Python 3.5.4 and 3.8.8?

We recommend that you install Content Moderation SDK for Python 3.x of mainstream versions. In this case, if the aliyunsdkcore library still fails to be installed, you can download the aliyunsdkcore library and import it to your project.

For more information, see Installation.

How do I install the aliyunsdkgreenextension utility class of Content Moderation SDK for Python?

You must download the aliyunsdkgreenextension utility class and import it to your project.

You must import the aliyunsdkgreenextension utility class to your project by using the following code:

from aliyunsdkgreenextension.request.extension import HttpContentHelper

For more information, see Installation.

Can a custom term library in Content Moderation contain terms in languages other than English?

No, a custom term library in Content Moderation can contain only letters and digits.

How long is the validity period of an OSS URL for an image or a video that is uploaded from a local machine to OSS for content moderation?

The validity period of an Object Storage Service (OSS) URL is 1 hour.

How am I charged for moderating a video live stream in Content Moderation?

The expense for moderating a video live stream depends on the number of frames captured from the video live stream. To calculate the number of captured frames, divide the duration of the video live stream by the time interval at which a frame is captured.

For example, the duration of a video live stream is 1 hour and a frame is captured every 5 seconds. In this case, the number of captured frames is calculated by using the following formula: 3,600 seconds/5 seconds = 720. Therefore, you are charged for the 720 captured frames.

How are the moderation results of a live stream returned to the callback URL in Content Moderation?

In Content Moderation, the moderation results of a video live stream are separately returned. Each time a violation is detected, a moderation result is returned. After the moderation of the live stream is complete, the overall moderation results are returned.

Why does my callback URL still receive data after Content Moderation stops moderating a live stream?

After the operation for moderating a live stream stops, the corresponding moderation task stops. However, your callback URL can still receive specific data because a low latency exists. If you have any questions, submit a .

Will a task for moderating a live stream stop if the live stream is interrupted or unavailable after the task is submitted?

After a task is submitted for moderating a live stream, no live stream is obtained because the live stream is interrupted or unavailable. In this case, Content Moderation requests the live stream for three times at specific intervals. The minimum interval is 10 seconds. If the live stream still fails to be obtained within 30 seconds, the task stops.

When will the status code 200 be returned during live stream moderation?

If the moderation is successful, the status code 200 is returned. if the live streaming is going on, the status code 280 is returned. If the live streaming ends or is interrupted, a corresponding status code is returned.

1.3. FAQ about OCR

This topic provides answers to the frequently asked questions (FAQ) about optical character recognition (OCR) in Content Moderation.

- How do I call an OCR operation to detect text in images in Content Moderation?
- What are the billing standards for Content Moderation OCR?

How do I call an OCR operation to detect text in images in Content Moderation?

You can call an OCR operation to submit OCR tasks to detect and obtain text in images. We recommend that you use Content Moderation SDK for Java to call the OCR operation to detect text in images or cards and certificates.

For more information, see Detect text in images synchronously and OCR.

What are the billing standards for Content Moderation OCR?

The fees of the images moderated per day are calculated by using the tiered pricing method.

For more information, see Pricing.

2.FAQ about billing

This topic provides answers to the frequently asked questions (FAQ) about Content Moderation billing.

- How am I charged for the Content Moderation service based on the pay-as-you-go billing method?
- Will the Content Moderation service be suspended if the service fee is overdue?

How am I charged for the Content Moderation service based on the pay-as-you-go billing method?

The Content Moderation API and the OSS violation detection feature support the pay-as-you-go billing method. The daily fee is settled based on the amount of data moderated on that day. For more information, see Daily fee calculation method.

For more information about the details of the pay-as-you-go billing method in Content Moderation, see Pricing.

Will the Content Moderation service be suspended if the service fee is overdue?

Alibaba Cloud generates a bill on each calendar day based on the number of resources that you used on the previous calendar day. Alibaba Cloud deducts the service fee accordingly from your Alibaba Cloud account. If you have purchased a subscription plan or pay-as-you-go plan, the resources used are first deducted from your resource plan. If the remaining resources in the plan are insufficient, the service fee is deducted from your account. If your account balance is insufficient to pay the bill and your account has an overdue payment for more than 24 hours, Alibaba Cloud stops providing the Content Moderation service for you.

For more information, see Overdue payment and renewal.

3.FAQ about error messages

This topic provides answers to the frequently asked questions (FAQ) about error messages that are returned after Content Moderation operations are called.

- Why is the following error message returned when I call a Content Moderation operation: You have not opened Yundun Content Moderation Service?
- Why is the following error message returned when I call a Content Moderation operation on an Elastic Compute Service (ECS) instance after the operation is called on my computer: SDK.ServerUnreachable: Speicified endpoint or uri is not valid?
- Why is the following error message returned when I call a Content Moderation operation on an Elastic Compute Service (ECS) instance after the operation is called on my computer: SDK.ServerUnreachable: Speicified endpoint or uri is not valid?
- Why am I prompted with the following message when I call a Content Moderation operation: algo failed(ocridcard-modelnot exist)?
- Why am I prompted with the following message when I call a Content Moderation operation: Your using subaccount is not authorized, please refer to the page?
- Why is the 401 (NOT_ALLOWED) error code returned when I call a Content Moderation operation?
- Why is the 400 (BAD_REQUEST) error code returned when I call a Content Moderation operation?
- Why is the 594 (EXPIRED) error code returned when I call a Content Moderation operation?
- Why is the 596 (PERMISSION_DENY) error code returned when I call a Content Moderation operation?
- Why is the AlgorithmTimeOut error returned or the detected text in the image incomplete when I call a synchronous OCR operation?
- Why is the 592 (DOWNLOAD_TIMEOUT) error code returned when I call a Content Moderation operation?
- Why is the "[task.datald] is too long(>256)" error message returned when I call a text moderation operation?
- Why is the "Status Code: 400" signature error returned when I send HTTP requests to call operations?
- Why is the 480 (DOWNLOAD FAILED) error code returned when I call a Content Moderation operation?
- Why is the 480 error code always returned for specific URLs when I call an operation to synchronously moderate videos?
- Why is the 480 (GIF_TOO_MUCH_PIXELS) error code returned when I call an operation to moderate images?
- Why is the 480 (Input/output error) error code returned when I call an operation to moderate video live streams?
- Why is the 500 (GENERAL_ERROR) error code returned when I call a Content Moderation operation?
- Why is the 500 (service interrupted) error code returned when I use Content Moderation SDKs?
- Why is the 586 (ALGO_FAILED) error code returned when I call a Content Moderation operation?
- Why is the 588 (EXCEED_QUOTA) error code returned when I call a Content Moderation operation?
- Why is the 586 error code returned when I call a Content Moderation operation to moderate images?
- Why is the "InvalidTimeStamp.Expired" error message returned when I call a Content Moderation operation?
- Why is an error reported when I use the human review feature in the Content Moderation console after the console is idle for a long period?

- Why is a string that contains HTML code tags moderated but no data returned when I use Content Moderation SDKs to moderate text?
- Why is the following error message returned when I call an OCR operation for content moderation before 09:00 in the morning and after 21:00 at night: Specified time stamp or date value is expired?
- Why is the 406 error code returned when I call a Content Moderation operation?
- Why is the following error message returned when I call an operation to manage the configurations of custom text libraries: The API is invalid?
- Why is the UnicodeEncodeError error returned when I run Content Moderation SDK for Python in Jupyter Notebook or by using a command-line tool?

For more information about the Content Moderation API, see FAQ about the Content Moderation API.

Why is the following error message returned when I call a Content Moderation operation: You have not opened Yundun Content Moderation Service?

Cause: Content Moderation is not activated.

Solution: Go to the activation page of Content Moderation to activate the service. Then, call the Content Moderation operation again.

Why is the following error message returned when I call a Content Moderation operation on an Elastic Compute Service (ECS) instance after the operation is called on my computer: SDK.ServerUnreachable: Speicified endpoint or uri is not valid?

Cause: The ECS instance cannot be accessed over the Internet. However, a Content Moderation operation must be called over the Internet.

Solution: Enable access to the ECS instance over the Internet. Check network connection by sending PING messages to the endpoint of the corresponding environment. If the network connection is normal, check whether the dependencies in a corresponding language are installed or the version of the **aliyun-**-sdk-core** library is as required. We recommend that you use the sample code provided in the SDK documentation to call Content Moderation operations.

For more information, see SDK overview.

Why is the following message returned when I use Content Moderation SDKs to call a Content Moderation operation on an ECS instance: SDK.ServerUnreachable : SocketTimeoutException has occurred on a socket read or accept?

Cause: A network access error occurred. A Content Moderation operation must be called over the Internet.

Solution: Enable access to the ECS instance over the Internet. To check network connection, you can send PING messages to the endpoint of the corresponding environment. If the network connection is normal, check whether the dependencies in a corresponding language are installed or the version of the **aliyun-**-sdk-core** library is as required. We recommend that you use the sample code provided in the SDK documentation to call Content Moderation operations.

For more information, see SDK overview.

Why am I prompted with the following message when I call a Content Moderation operation: algo failed(ocridcardmodelnotexist)?

Cause: The optical character recognition (OCR) card model cannot be used in the region that corresponds to the current endpoint.

Solution: Select the endpoint of a region where the OCR card model can be used.

For more information, see Endpoint.

Why am I prompted with the following message when I call a Content Moderation operation: Your using subaccount is not authorized, please refer to the page?

Cause: The RAM user is not authorized to call Content Moderation operations.

Solution: You can call a Content Moderation operation as a RAM user. Before that, you must create a RAM user and authorize the RAM user to call Content Moderation operations.

For more information, see Call the Content Moderation API as a RAM user.

Why is the 401 (NOT_ALLOWED) error code returned when I call a Content Moderation operation?

Cause: The image, video URL that you use is not secure.

Solution: Check whether your URL is secure or whether your IP address belongs to the following internal CIDR blocks. You must use a public IP address to access Content Moderation.

- 10.0.0/8
- 11.0.0.0/8
- 100.64.0.0/10
- 172.16.0.0/12
- 192.168.0.0/16
- 127.0.0.1/32
- 33.0.0.0/8

Why is the 400 (BAD_REQUEST) error code returned when I call a Content Moderation operation?

Cause: A request parameter is invalid. If the **[task.url] is bad format** error message is returned, the URL is invalid.

Solution: We recommend that you check whether the request parameters of the Content Moderation operation are specified as required. If you are prompted with a message that indicates an invalid URL, check the URL. If you can open the URL in a browser, check whether the URL contains special characters. If the URL contains special characters, encode the special characters.

For more information, see Common parameters.

Why is the 594 (EXPIRED) error code returned when I call a Content Moderation operation?

Cause: The task ID expired. For example, you can call an operation to query the moderation results of an asynchronous task of image moderation within 24 hours after the task is run. After 24 hours, the **EXPIRED** message is returned.

Solution: After you submit the asynchronous task of image moderation, we recommend that you set the interval for querying moderation results to 30 seconds. This ensures that the task ID does not expire.

Why is the 596 (PERMISSION_DENY) error code returned when I call a Content Moderation operation?

Cause: Content Moderation is not activated, or the account has no permissions, is disabled, or has an overdue payment. For more information, see the returned error message.

Solution: Check whether the account that you use to call the Content Moderation operation has permissions or has an overdue payment. If you call a Content Moderation operation as a RAM user, you must grant permissions to the RAM user. If you have not activated Content Moderation, log on to the activation page of Content Moderation to activate the service.

For more information, see Call the Content Moderation API as a RAM user.

Why is the AlgorithmTimeOut error returned or the detected text in the image incomplete when I call a synchronous OCR operation?

Cause: The **AlgorithmTimeOut** error indicates a moderation timeout. Image OCR is time-consuming. The default timeout period for a synchronous operation is 3 seconds. If you call a synchronous OCR operation to detect a large amount of text in an image, a timeout error is very likely to occur.

Solution: If your business requires you to process images that contain a large amount of text, we recommend that you call the /green/image/asyncscan operation. This way, Content Moderation retries the OCR operation multiple times and optimizes the OCR process to avoid a timeout error.

Why is the 592 (DOWNLOAD_TIMEOUT) error code returned when I call a Content Moderation operation?

Cause: The download timed out. The download duration is limited to 3 seconds. The size of the content to be moderated must be allowed by the Content Moderation operation.

Solution: We recommend that you check whether the image URL is accessible or whether the image can be downloaded within 3 seconds. If an Alibaba Cloud CDN (CDN) domain is used, the download timeout may be caused due to the communication between CDN and the origin server. Therefore, we recommend that you do not use a CDN domain. If you access services across regions, a timeout error is very likely to occur. For example, you want to access an Object Storage Service (OSS) bucket in the US (Silicon Valley) region from the Singapore (Singapore) region. We recommend that you access an OSS bucket in the same region.

Why is the "[task.dataId] is too long(>256)" error message returned when I call a text moderation operation?

Cause: The data ID exceeded 128 characters in length. A data ID uniquely identifies business data. A data ID can contain letters, digits, underscores (_), hyphens (-), and periods (.).Data ID example: cfd33235-71a4-468b-8137-a5ffe323a7e8.

Solution: Specify the dataid parameter in the API request based on the API reference.

For more information, see Text Synchronous moderation.

Why is the "Status Code: 400" signature error returned when I send HTTP requests to call operations?

Cause: The signature mechanism for HTTP requests is complex. If you construct a signature on your own, code errors are likely to occur.

Solution: We recommend that you use Content Moderation SDKs. The signature process is encapsulated in the SDKs, and you do not need to write the code for signature. If you can call operations only by using HTTP requests, we recommend that you first understand the signature mechanism of Content Moderation.

For more information, see SDK overview and Signature method.

Why is the 480 (DOWNLOAD_FAILED) error code returned when I call a Content Moderation operation?

Cause: The download failed. This may be because the URL of the object to be moderated is inaccessible or the content to be moderated cannot be downloaded. This may also be because the size or resolution of the content to be moderated exceeds the upper limit.

Solution: If the object is an image, we recommend that you check whether the image URL is accessible. If the object is a video, specific images in the sequence of captured frames may fail to be downloaded. In this case, we recommend that you check whether frame-related parameters, such as the time parameter, are valid. If the object is a video stream, we recommend that you check whether the video stream is being pushed during the moderation. You can also submit an OSS ticket to troubleshoot the error.

Why is the 480 error code always returned for specific URLs when I call an operation to synchronously moderate videos?

Cause: The download failed. This may be because the URL of the object to be moderated is inaccessible or the content to be moderated cannot be downloaded. This may also be because the size or resolution of the content to be moderated exceeds the upper limit.

Solution: If the object is a video, specific images in the sequence of captured frames may fail to be downloaded. In this case, we recommend that you check whether frame-related parameters, such as the time parameter, are valid. You can also submit an OSS ticket to troubleshoot the error.

For more information, see Video snapshots.

Why is the 480 (GIF_TOO_MUCH_PIXELS) error code returned when I call an operation to moderate images?

Cause: The download failed. The **GIF_TOO_MUCH_PIXELS** error message indicates that the size or resolution of the content to be moderated exceeds the upper limit.

Solution: Content Moderation can moderate an image of which the size does not exceed 20 MB, the height or width does not exceed 30,000 pixels, and the total resolution does not exceed 0.25 billion pixels. If the size or resolution of an image to be moderated exceeds the upper limit, compress the image and then submit it for moderation.

For more information, see Synchronous moderation.

Why is the 480 (Input/output error) error code returned when I call an operation to moderate video live streams?

Cause: The download failed. In general, the **Input / output error** error message is returned because an error occurred in media streaming files. To be specific, the error message is returned because the URL of the content to be moderated is inaccessible or the content cannot be downloaded.

Solution: If the object is a video stream, we recommend that you check whether the video stream is being pushed and whether the live streaming is stopped. If the object is a video, specific images in the sequence of captured frames may fail to be downloaded. In this case, we recommend that you check whether frame-related parameters, such as the time parameter, are valid.

Why is the 500 (GENERAL_ERROR) error code returned when I call a Content Moderation operation?

Cause: A temporary error occurred on the server. Troubleshoot the error based on the returned error message.

Solution: If the error code is occasionally returned, try again. If the error code is continuously returned, submit a ticket and provide the returned error information to help technical engineers troubleshoot the error at the earliest opportunity.

Why is the 500 (service interrupted) error code returned when I use Content Moderation SDKs?

Cause: A temporary error occurred on the server. The **service interrupted** error message may be returned because your SDK version is incompatible with the server version.

Solution: We recommend that you check whether the latest SDK version is used. If the SDK version is not the latest, update the SDK.

For more information, see Installation.

Why is the 586 (ALGO_FAILED) error code returned when I call a Content Moderation operation?

Cause: An error occurred in the algorithm service. In general, the error code is returned because a timeout error occurred due to the network jitter of the algorithm service.

Solution: If the error code is occasionally returned, try again. If this error code is continuously returned, submit a ticket and provide the returned error information to help technical engineers troubleshoot the error at the earliest opport unity.

Why is the 588 (EXCEED_QUOTA) error code returned when I call a Content Moderation operation?

Cause: The number of moderation requests per second exceeded the concurrency limit. By default, a maximum of 50 images, 20 videos, or 100 text entries can be concurrently moderated per second.

Solution: We recommend that you reduce the number of moderation requests per second and queue the content to be moderated. Alternatively, you can contact your customer manager to raise the concurrency limit based on your concurrency requirements during peak hours.

Why is the 586 error code returned when I call a Content Moderation operation to moderate images?

Cause: An error occurred in the algorithm service. If the error code is returned only for specific images or videos, a codec failure may occur.

For more information, see Synchronous moderation.

image formats are supported: PNG, JPG, JPEG, BMP, GIF, and WEBP.

Why is the "InvalidTimeStamp.Expired" error message returned when I call a Content Moderation operation?

Cause: The timestamp for calling the operation expired. This error message is returned because the offset between the timestamps on the client and server exceeds 15 minutes.

Solution: Check whether the server time is recently adjusted or whether the time zone on the client is the same as that on the server.

For more information, see Common parameters.

Why is an error reported when I use the human review feature in the Content Moderation console after the console is idle for a long period?

Cause: This is a console access error. It may be caused due to different reasons. For example, it may be caused because the access authorization failed due to an invalid logon session. Solution: If the **Request ID** error is returned,

we recommend that you refresh your browser and log on to the console again. If an error is continuously reported after a successful logon, submit a ticket and provide the error information or error screenshots. This helps technical engineers troubleshoot the error at the earliest opportunity.

Why is a string that contains HTML code tags moderated but no data returned when I use Content Moderation SDKs to moderate text?

Cause: If the string to be moderated contains code tags such as HTML code tags, the firewall determines that the string contains risky code and blocks your network request. As a result, no response is returned for the request.

Solution: We recommend that you filter out code tags such as the characters that contain HTML code or SQL code in text before you moderate the text. This prevents the text from being blocked by the firewall.

For more information, see Text Synchronous moderation.

Why is the following error message returned when I call an OCR operation for content moderation before 09:00 in the morning and after 21:00 at night: Specified time stamp or date value is expired?

Cause: The timestamp for calling the operation expired. This error message is returned because the offset between the timestamps on the client and server exceeds 15 minutes.

Solution: We recommend that you use the SDKs provided by the Alibaba Cloud official website to call the operation. Content Moderation supports SDKs for different programming languages such as Java, Python, and PHP. If you can use only HTTP requests to call the operation, check whether the server time is recently adjusted or whether the time zone on the client is the same as that on the server.

For more information, see Common parameters.

Why is the 406 error code returned when I call a Content Moderation operation?

Cause: This may be because the used and referenced SDK versions are not the same. For example, you use the code of a later SDK version to call operations but reference an earlier SDK version.

Solution: We recommend that you check whether the referenced SDK version is the latest. If you use Content Moderation SDK for Java, update the referenced core and green packages to the latest version. Content Moderation supports SDKs for different programming languages such as Java, Python, and PHP.

For more information, see SDK overview.

Why is the following error message returned when I call an operation to manage the configurations of custom text libraries: The API is invalid?

Cause: The operation that is used to manage the configurations of custom text libraries can be called only in the China (Shanghai) region. You cannot call this operation in another region such as the China (Beijing) region. You can call this operation to query the text libraries of a specific region. If no text library is created in the specific region, no data is returned after you call this operation.

Solution: Call the operation that is used to manage the configurations of custom text libraries in the China (Shanghai) region.

For more information, see Create a text library.

Why is the UnicodeEncodeError error returned when I run Content Moderation SDK for Python in Jupyter Notebook or by using a command-line tool?

Cause: An error occurred in the integrated development environment (IDE).

Solution: We recommend that you run Content Moderation SDK for Python in a mainstream IDE such as PyCharm. The IDE error occurred in the specific IDE is unknown. We recommend that you check the IDE to troubleshoot the error.

4.FAQ about moderation effects

This topic provides answers to the frequently asked questions (FAQ) about the moderation effects of the Content Moderation API.

- Why is normal content mistaken as abusive content during text moderation? Why does abusive content fail to be detected during text moderation?
- Why is normal content mistaken as pornographic content during text moderation? Why does pornographic content fail to be detected during text moderation?
- Why do ads such as QR codes fail to be detected during text moderation?
- Why do pornographic images fail to be detected during image moderation?
- Why are the terms added to the ignore list still blocked during text moderation?
- Why are the images that contain national flags and emblems not blocked during image moderation?
- What do I do if a specific part of a human body in a medical image is detected as pornographic content in Content Moderation?

Why is normal content mistaken as abusive content during text moderation? Why does abusive content fail to be detected during text moderation?

Abusive content in text can be classified into serious abuse, mild abuse, and colloquialism based on the abusive severity. You can modify the policy for machine-assisted text moderation for your business scenario in the Content Moderation console. If the abusive content in specific text fails to be detected or the normal content is mistaken as abusive content, we recommend that you create a custom text pattern or term library. Then, specify an ignore list or a review list to ignore or review specific terms.

For more information, see Customize policies for machine-assisted moderation and Manage custom text libraries.

Why is normal content mistaken as pornographic content during text moderation? Why does pornographic content fail to be detected during text moderation?

Pornographic content in text can be classified into serious pornography, vulgar content, and sexual knowledge based on the pornographic severity. You can modify the policy for machine-assisted text moderation for your business scenario in the Content Moderation console. If the pornographic content in specific text fails to be detected or the normal content is mistaken as pornographic content, we recommend that you create a custom text pattern or term library. Then, specify an ignore list or a review list to ignore or review specific terms.

For more information, see Customize policies for machine-assisted moderation and Manage custom text libraries.

Why do ads such as QR codes fail to be detected during text moderation?

Ads in text are classified into phone numbers, WeChat accounts, QQ accounts, URLs, and slogans. You can modify the policy for machine-assisted text moderation for your business scenario in the Content Moderation console. If the ads in specific text fail to be detected, we recommend that you create a custom text pattern or term library. Then, specify a blacklist or a review list to block or review specific terms.

For more information, see Customize policies for machine-assisted moderation and Manage custom text libraries.

Why do pornographic images fail to be detected during image moderation?

Pornographic images can be classified into purely pornographic, vulgar and indecent, and sexy images based on the sexual explicitness. You can modify the policy for machine-assisted image moderation for your business scenario in the Content Moderation console. If specific pornographic images fail to be detected, we recommend that you create a custom image library and select a scenario for detecting pornographic content in images. Then, specify a blacklist or a review list to block or review the specific pornographic content in images.

For more information, see Customize policies for machine-assisted moderation and Manage custom text libraries.

Why are the terms added to the ignore list still blocked during text moderation?

The terms added to a custom text library may not take effect due to the following reasons:

- New terms take about 15 minutes to be effective after they are added to a custom text library. We recommend that you try again later.
- Make sure that the text type and matching method are valid.
 - We recommend that you manage the content that contains less than or equal to five characters as terms and use fuzzy match.
 - We recommend that you manage the content that contains more than five characters as text patterns and use exact match.
- The custom text library must be applied to the corresponding business scenario before it takes effect.

For more information, see Manage custom text libraries.

Why are the images that contain national flags and emblems not blocked during image moderation?

Content Moderation provides the following image moderation scenarios: pornography detection, terrorist content detection, ad violation detection, and undesirable scene detection. If you want to block images that contain national flags and emblems, select the scenario of terrorist content detection. Then, modify the policy for machine-assisted image moderation for your business scenario in the Content Moderation console. In the scenario of terrorist content detection, the following types of images can be detected: figure, symbol, ordnance, incident, religion, public service, and ticket. In this case, select **The National Flag and Emblem of the People's Republic of China** for Symbol Recognition. In addition, you must set the scenes parameter to terrorism in the moderation request.

For more information, see Customize policies for machine-assisted moderation and Synchronous moderation.

What do I do if a specific part of a human body in a medical image is detected as pornographic content in Content Moderation?

Whether an image is a medical image cannot be identified during image moderation for pornographic content. We recommend that you set the dataid parameter in the moderation request to mark medical images. This way, human review can be further performed on image moderation results.

For more information, see Synchronous moderation.