

图片内容安全 快速入门



腾讯云

【版权声明】

©2013–2025 腾讯云版权所有

本文档（含所有文字、数据、图片等内容）完整的著作权归腾讯云计算（北京）有限责任公司单独所有，未经腾讯云事先明确书面许可，任何主体不得以任何形式复制、修改、使用、抄袭、传播本文档全部或部分内容。前述行为构成对腾讯云著作权的侵犯，腾讯云将依法采取措施追究法律责任。

【商标声明】



及其它腾讯云服务相关的商标均为腾讯云计算（北京）有限责任公司及其关联公司所有。本文档涉及的第三方主体的商标，依法由权利人所有。未经腾讯云及有关权利人书面许可，任何主体不得以任何方式对前述商标进行使用、复制、修改、传播、抄录等行为，否则将构成对腾讯云及有关权利人商标权的侵犯，腾讯云将依法采取措施追究法律责任。

【服务声明】

本文档意在向您介绍腾讯云全部或部分产品、服务的当时的相关概况，部分产品、服务的内容可能不时有所调整。

您所购买的腾讯云产品、服务的种类、服务标准等应由您与腾讯云之间的商业合同约定，除非双方另有约定，否则，腾讯云对本文档内容不做任何明示或默示的承诺或保证。

【联系我们】

我们致力于为您提供个性化的售前购买咨询服务，及相应的技术售后服务，任何问题请联系 4009100100或95716。

文档目录

快速入门

图片内容安全服务

图片 AI 生成识别服务

快速入门

图片内容安全服务

最近更新时间：2025-08-19 16:52:41

步骤一：登录注册

注册腾讯云账号，并完成 [实名认证](#)，详情请参见 [账号注册教程](#)。

- 说明：**
如果您已拥有腾讯云账号，可略过此步骤。

步骤二：创建子账号（可选）

在 [注册腾讯云](#) 账号时，所生成的账号为主账号，拥有该主账号下所有云资源的管理权限。如果您需要其他用户来协助您一起管理账号下的云资源，可以通过使用访问管理功能创建子账号，并为其绑定权限策略。该功能可以创建一个或多个子账号提供给团队成员，详情请参见 [新建子用户](#)。
子账号刚创建时，默认是没有任何权限的，需要主账号或管理员为其绑定策略，才能使子账号拥有某些云资源的操作权限。您可以为子账号配置内容安全 CAM 授权，详情请参见 [CAM 授权指引](#)。

步骤三：开通服务

前提条件

- 已获取腾讯云主账号或子账号。
- 登录账号已拥有图片内容安全服务的读写权限。
- 账户余额需大于0元，如账户余额小于0元，请完成 [账户充值](#) 后前往开通服务。

试用说明

- 新用户免费试用权益：**首次开通内容安全服务的用户可免费试用3000张图片（不含“图片 AI 生成识别”能力），试用套餐包有效期为15天。

- 说明：**
首次开通领取试用包仅为获取测试权益，非正式服务开通。

- 后续计费与服务开通规则**
 - 若试用包到期未用完、提前用完，需 [购买正式服务](#) 或开通后付费服务，否则无法正常调用接口。
 - 后付费模式结算方式为日结，单价标准为25元 / 万张图片。
- 常见问题提示**
 - 老用户（曾领取过试用包）接入时，若未开通正式服务，会提示“服务未开通”。
 - 新用户领取试用包后，若15天有效期内未使用或未开通正式服务，试用包过期后接入也会提示“服务未开通”。
- 解决方案：**若遇到“服务未开通”提示，需 [购买正式服务](#) 或开通按量后付费功能，完成后即可正常调通接口。



操作步骤

1. 登录 [内容安全控制台](#)，单击立即开通。



2. 选择图片内容安全服务，单击确定完成开通。



步骤四：配置策略

建议您使用配置任务策略，可根据业务需求配置识别策略，用于个性化服务体验。

说明

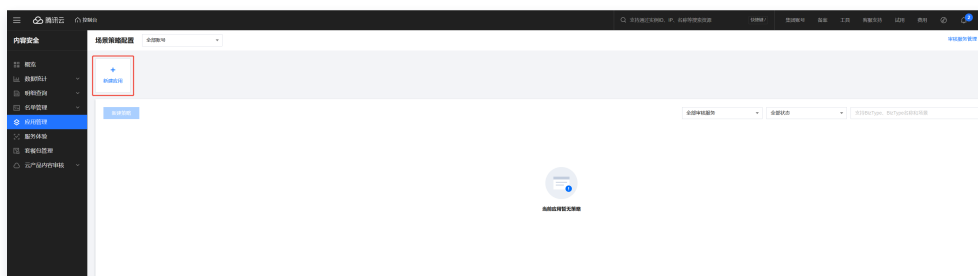
- 腾讯云内容安全服务已预设有默认策略，如使用默认策略，可略过此步骤。
- 默认策略为天御多行业模型沉淀的策略配置，适用于大部分的内容安全需求。
- 如您想要配置更符合业务场景的策略，可按照以下步骤创建自定义策略实现个性化服务。

前提条件

- 已获取腾讯云主账号或子账号。
- 登录账号已拥有图片内容安全服务的读写权限。
- 登录账号已开通图片内容安全服务。

操作步骤

1. 登录 [内容安全控制台](#)，在概览页单击应用管理 > 新建应用，新建需要配置文本内容安全管理的应用。



2. 在应用管理页面，单击新建应用，输入应用名称，单击保存完成应用创建。

新建应用

✕

应用名称

测试_图片

仅支持数字、中英文、下划线，最多20个字符

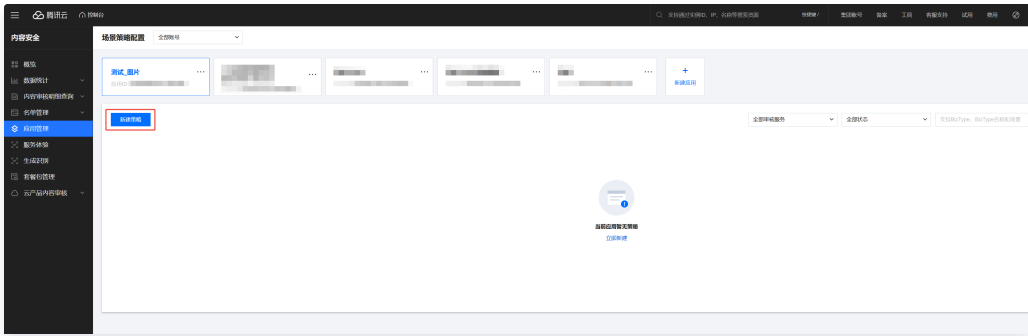
应用说明

非必填项

保存

取消

3. 在应用管理页面，单击新建策略。



4. 在新建策略页面，选择图片内容安全，输入Biztype、Biztype名称、场景，单击下一步进入策略模板配置。

新建策略

✕

1 策略基础信息

>

2 策略模板配置

管理审核服务

☐ 文本内容安全

☒ 图片内容安全

☐ 音频内容安全

☐ 视频内容安全

☐ 视频流内容安全

☐ 音频流内容安全

管理审核服务

Biztype

text_ims1

BizType 是策略集，一个业务场景要对应到一个BizType。
例:biztype='nickname_input', biztype名称='昵称审核', 场景='个人主页'。

Biztype名称

测试_图片

场景

非必填项，用于对BizType的分类

下一步

取消

参数名称	描述
场景名称	场景的文字描述，可使用中文、英文、数字及下划线组合，长度不超过20个字符
关联审核服务	选择场景需要关联的审核服务

5. 在策略模板配置页面，根据具体业务选择对应的行业、业务类型、业务场景，并单击确定保存，即可完成策略创建。

6. 策略创建完成，在策略详情页面会显示当前系统使用的策略，如需自定义更改，单击**策略配置**，进入策略详情，选择您想开启的识别策略。

版权所有：腾讯云计算（北京）有限责任公司

图片审核策略详情

场景名称: 测试-图片

1 策略基础配置

>

2 关联名单库

策略基础信息

* 策略名称

测试-图片

* bizType名称

text_jms1

识别策略配置

策略模板

①

请基于业务需求选择需要的标签分类, 选中对应分类代表需要进行识别打击该内容

三级标签是模型用于过滤, 系统命中后会返回二级标签

☑ 图片图像识别

☒ 图片文本识别

[-] 人脸

4

恢复默认

标签名称	标签说明	标签级别	检测状态
色情	8	二级	☐
性感	0	二级	☐
暴恐	1	二级	☐
实体	3	二级	☒
台标	0	二级	☐
二维码	0	二级	☐
地图	4	二级	☒
广告	0	二级	☒
涉政建筑	4	二级	☒

[-] 色情

[-] 性感

[-] 暴恐

[-] 实体

[-] 台标

[-] 二维码

☒ 地图

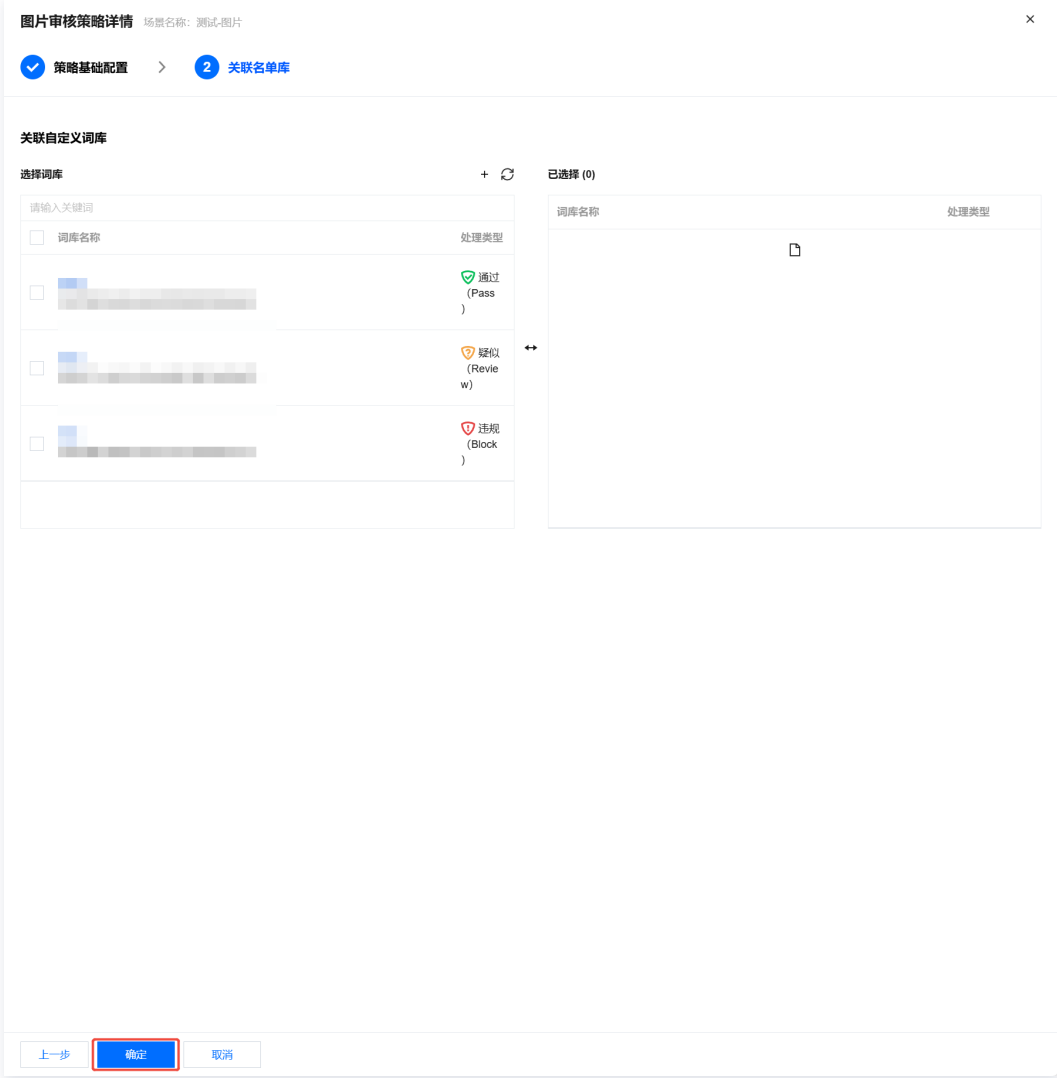
[-] 广告

[-] 涉政建筑

下一步

取消

7. 完成识别策略配置后, 单击下一步, 选择是否需要关联名单库用于内容识别, 如暂无关联名单库, 可跳过, 或保存当前策略后前往。



8. 完成后单击**确认**，完成策略配置。

参数名称	描述
识别策略配置	根据业务需求，选择是否需要识别不同类型的内容。
关联自定义库	选择是否需要使用自定义词库用于内容识别，如暂无自定义库，可跳过，或保存当前策略后前往 步骤五：配置自定义词库 。

步骤五：配置词库（可选）

前提条件

- 已获取腾讯云主账号或子账号。
- 登录账号已拥有图片内容安全服务的读写权限。
- 登录账号已开通图片内容安全服务。

预设词库

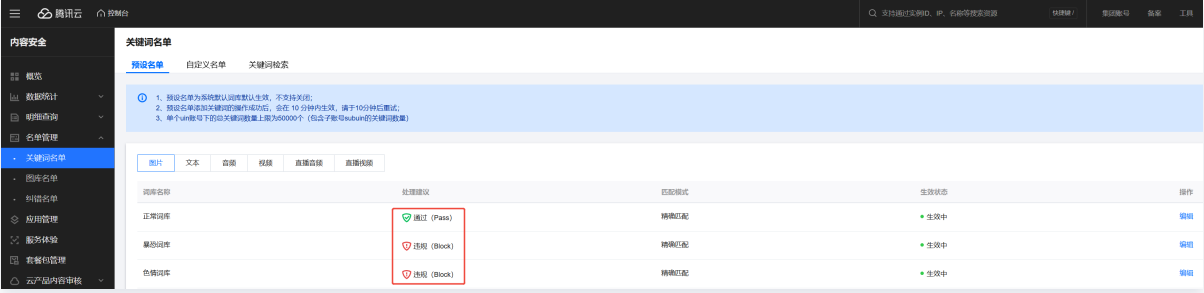
说明

如无需添加预设词库关键词，可略过此步骤。

1. 登录 [内容安全控制台](#)，在左侧导航栏中，单击**名单管理 > 关键词名单**。
2. 在预设名单页面，产品类型选择**图片**。



3. 在预设名单页面，支持查看系统预设的词库，包括：正常词库、违法词库、广告词库等。



说明：
预设词库的不同颜色代表不同的屏蔽逻辑，绿色代表通过，红色代表违规。

4. 在预设名单页面，单击编辑，可以添加关键词。

说明：
预设词库默认开启，不支持关闭、删除操作。

5. 在编辑词库页面，单击添加关键词，输入关键词，单击确定，即可保存关键词至当前词库。

← 管理词库

通过

正常

完成添加的关键词10分钟内生效

匹配模式

精确匹配

状态

● 生效中

添加关键词

批量删除

☐ 关键词

样本ID

机审标签

关键词类型

添加关键词

×

关键词类型

☒ 普通词

☐ 组合词

☐ 英文

添加关键词 *

多个以回车换行分隔，一次提交不超过2000个

关键词备注

请输入关键词的备注说明，最多可输入100个字

0 / 100

1、每个关键词以换行来确定，单个关键词长度20 个汉字以内

2、支持以复制粘贴的方式批量导入，单次最多添加关键词2000个，词与词之间用换行符分隔

3、单个uin账号下的总关键词数量上限为50000个（包含子账号subuin的关键词数量）

4、如需添加组合词，选择关键词类型：组合词，使用&符号对两个或者多个词进行连接，添加格式为“关键词1&关键词2...”，其中关键词只支持纯中文，这些关键词同时出现时则会命中

确定

取消

参数名称	描述
关键词	- 每个关键词以换行来确定，单个关键词长度20 个汉字以内。 - 支持以复制粘贴的方式批量导入，单次最多添加关键词2000个，词与词之间用换行符分隔。 - 单个词库的关键词添加个数上限为10000个。 - 单个 uin 账号下的总关键词数量上限为50000个（包含子账号 subuin 的关键词数量）。 - 如需添加组合词，选择关键词类型：组合词，使用&符号对两个或者多个词进行连接，添加格式为“关键词1&关键词2...”，其中关键词只支持纯中文，这些关键词同时出现时则会命中。

自定义词库

自定义词库用于添加个性化的识别词库，识别图片上是否含有违规文本。

ⓘ 说明

如无需配置自定义词库，可略过此步骤。

1. 登录 [内容安全控制台](#)，在左侧导航栏中，选择**名单管理 > 关键词名单 > 自定义名单**。

2. 在自定义名单页面，单击**新建词库**，填写文本库名称及按业务需求添加关键词。

新建关键词库

词库名称

请输入关键词库名称

可使用中文、英文、数字及下划线组合，长度32个字符内

识别结果

☒ 违规 (Block)

☐ 疑似 (Review)

☐ 放过 (Pass)

匹配模式

☒ 模糊匹配

☐ 精确匹配

匹配模式选择后，不可变更。匹配模式的逻辑分别为：
精确匹配：对输入的文本和关键词做直接的一致性匹配，例如：配置类关键词“香蕉”，则“我想吃香蕉”会命中，但“我想吃香x蕉”则不命中
模糊匹配：用于检测可能发生变体的词，当前支持拆分字、形似字、音似字、简繁体、大小写、大写数字六种形式。

确定

取消

参数名称	描述
词库名称	词库的文字描述，可使用中文、英文、数字及下划线组合，长度不超过32个字符
处理建议	可选择违规、疑似、放过： - 违规：确认为不良信息 - 疑似：可能为不良信息，需要人工识别 - 放过：业务方认为的正常信息。
匹配模式	可选择精确匹配或模糊匹配： - 精确匹配：对输入文本进行匹配，匹配对象需完全一致 - 模糊匹配：用于检测可能发生变体的词，当前支持拆分字、形似字、音似字、简繁体、大小写、大写数字等形式的相似词进行匹配

3. 单击**确定**，即可创建自定义词库。

4. 在自定义词库页面下方的列表中，将显示刚创建的词库。在创建词库弹窗，填写自定义库的相关信息，结合业务需求配置自定义库。

内容安全

关键词名单

预设名单

自定义名单

关键词状态

1. 对于名单内关键词操作成功时，会在 10 分钟内生效
2. 每个关键词下最多支持设置最多 1000000 个（包含子账号 subaccount 的关键词数量）

新增词库

输入词库名称

词库名称	识别结果	匹配模式	最后修改时间	操作
测试0	通过 (Pass)	模糊匹配	2025-03-26 11:51:15	关联场景 管理 删除 导出
测试1	疑似 (Review)	模糊匹配	2025-03-26 11:51:08	关联场景 管理 删除 导出
测试2	违规 (Block)	模糊匹配	2025-03-26 11:50:19	关联场景 管理 删除 导出

说明：
自定义词库的不同颜色代表不同的屏蔽逻辑，红色代表违规，橙色代表疑似，绿色代表正常。

5. 在自定义词库页面，单击**管理**，选择添加关键词，单击**确定**，即可保存关键词至当前词库。

添加关键词

机审标签

正常

关键词类型

☒ 普通词

☐ 组词

☐ 英文

添加关键词

输入框添加

文件上传

输入框一次最多添加2000个词。若超过该数量，则请使用文件上传的方式。

请输入需要添加的关键词，每个关键词以换行来确定，单个关键词长度20个汉字以内，添加多个关键词以回车换行分隔，一次提交不超过2000个

关键词备注

请输入关键词的备注说明，最多可输入100个字

0 / 100

确定

取消

参数名称	描述
处理建议	识别模型对应的违规类型。
关键词	- 每个关键词以换行来确定，单个关键词长度20 个汉字以内。 - 支持以复制粘贴的方式批量导入，单次最多添加关键词2000个，词与词之间用换行符分隔。 - 单个 uin 账号下的总关键词数量上限为50000个（包含子账号 subuin 的关键词数量）。 - 如需添加组词，选择关键词类型：组词，使用&符号对两个或者多个词进行连接，添加格式为“关键词1&关键词 2...”，其中关键词只支持纯中文，这些关键词同时出现时则会命中。

说明：

自定义词库配置完成后，可以在 [步骤四：配置策略](#) 时将自定义词库与策略关联使用。

步骤六：配置图库（可选）

前提条件

- 已获取腾讯云主账号或子账号。
- 登录账号已拥有图片内容安全服务的读写权限。
- 登录账号已开通图片内容安全服务。

预设图库

说明

如无需添加预设图库图片，可略过此步骤。

- 登录 [内容安全控制台](#)，在左侧导航栏中，单击[名单管理](#) > [图库名单](#)。
- 在图库名单页面，产品类型选择[图片](#)。



3. 在图库名单页面，支持查看系统预设的图库，包括：正常图片库、暴恐图片库、广告图片库等。

图片 视频 直播视频				
图库名称	识别结果	匹配模式	生效状态	操作
正常图片库	通过 (Pass)	精确匹配	生效中	编辑
暴恐图片库	违规 (Block)	精确匹配	生效中	编辑

说明：
预设图库的不同颜色代表不同的屏蔽逻辑，绿色代表通过，红色代表违规。

4. 在图库名单页面，单击**编辑**，可以添加图片。

说明：
预设图库默认开启，不支持关闭、删除操作。

5. 在编辑图库页面，单击**添加样本**，点击上传图片，单击**确定**，即可保存样本图片至图库。

**说明：**

- 预设图库样本图片格式支持 jpg, jpeg, png, webp。
- 单次上传支持50张。
- 预设图库默认容量为10000张。

步骤七：体验服务

完成以上步骤后，您可以选择创建的识别策略，通过立即识别图片，体验图片内容安全服务。

前提条件

- 已获取腾讯云主账号或子账号。
- 登录账号已拥有图片内容安全服务的读写权限。
- 登录账号已开通图片内容安全服务。

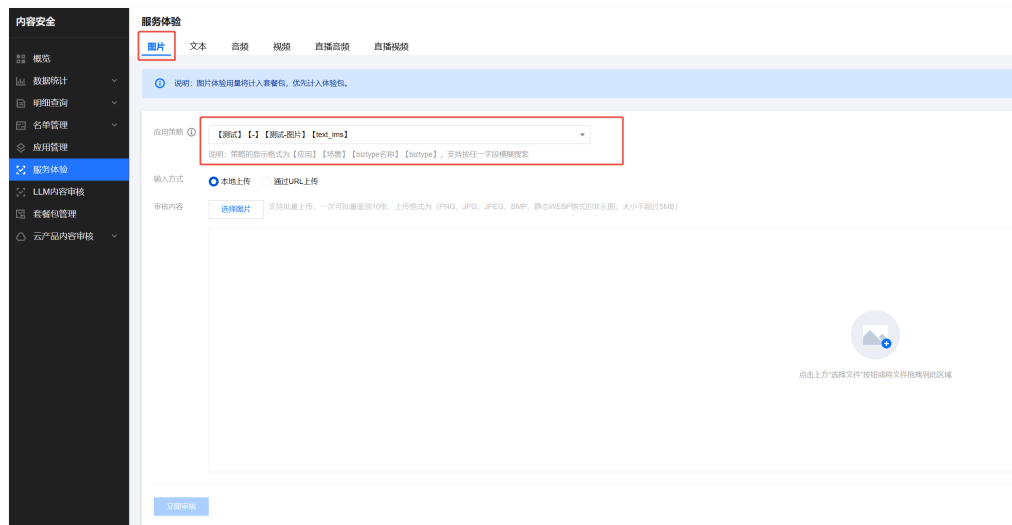
操作步骤

1. 登录 [内容安全控制台](#)，在左侧导航栏中，选择**服务体验 > 图片**。
2. 在服务体验页面，选择所需的应用策略，通过以下任意一种方式指定识别图片：
 - 在识别内容输入框内填入识别内容 URL 或 COS 访问地址，单击**添加网络图片**。
 - 单击**选择文件**，直接上传本地图片。

说明：

图片文件需满足以下要求：

- 图片文件格式：PNG、JPG、JPEG、BMP、GIF、静态WEBP。
- 图片文件大小：5 MB以内。
- 图片文件分辨率：建议分辨率大于256×256，否则可能会影响识别效果。
- 图片文件链接支持的传输协议：HTTP、HTTPS。



3. 单击立即审核，在图片下方会显示该图片的识别结果。

步骤八：接入服务

如果您需要接入图片内容安全服务，需通过调用 API 接口完成接入，详情请参考 [API 接入指引](#)。

说明

在调用 API 接口之前，您需要获取腾讯云 API 访问密钥。腾讯云通过 SecretId 和 SecretKey 对开发者的身份和权限进行校验，您可以参考以下步骤获取腾讯云 API 访问密钥：

进入 [云 API 密钥管理](#) 页面，在左侧导航栏选择访问密钥 > API 密钥管理，单击新建密钥创建密钥，保存 SecretId 和 SecretKey 供后续 API 调用传参。

图片 AI 生成识别服务

最近更新时间：2025-06-03 17:49:01

服务概述

图片内容安全提供了对 AI 生成内容的识别能力，本文将为您介绍[页面体验](#)和 **API 调用**两种图片生成识别服务的使用方式。

快捷入口：

- 页面体验：[内容安全控制台-生成识别](#)
- API 调用：[图片内容安全接入文档](#)

能力介绍

大模型技术的快速发展，给内容创作领域带来了更多的可能性，但同时也降低了创作门槛，带来了虚假信息、生成非法内容等问题。本产品基于图像域和频域的双重模态信息，通过风格识别、频谱分析等技术手段，挖掘图像的AIGC生成痕迹，无论是油画风格、拟态风格还是艺术风格，都能精准识别是否为 AI 生成。识别范畴上，已覆盖 Stable Diffusion 、Midjourney 、Dalle3 、GPT-4o 等主流大模型，在实际应用中识别准召率高。

服务对接

说明：

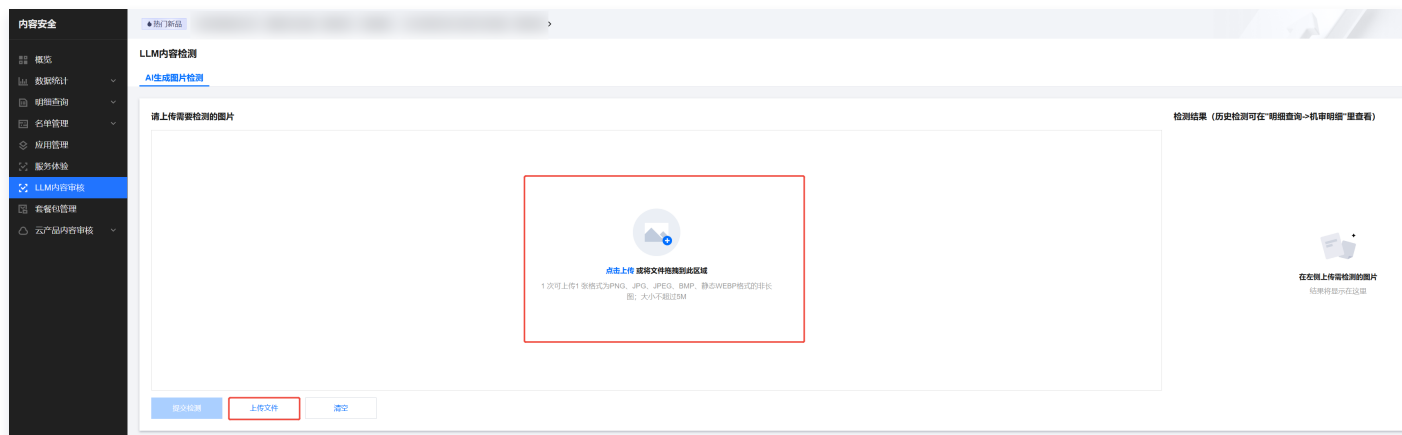
本服务的使用，需要您先登录腾讯云账号。若您未注册腾讯云账号，可参考 [账号注册教程](#) 先行注册。

方式一 通过控制台页面使用

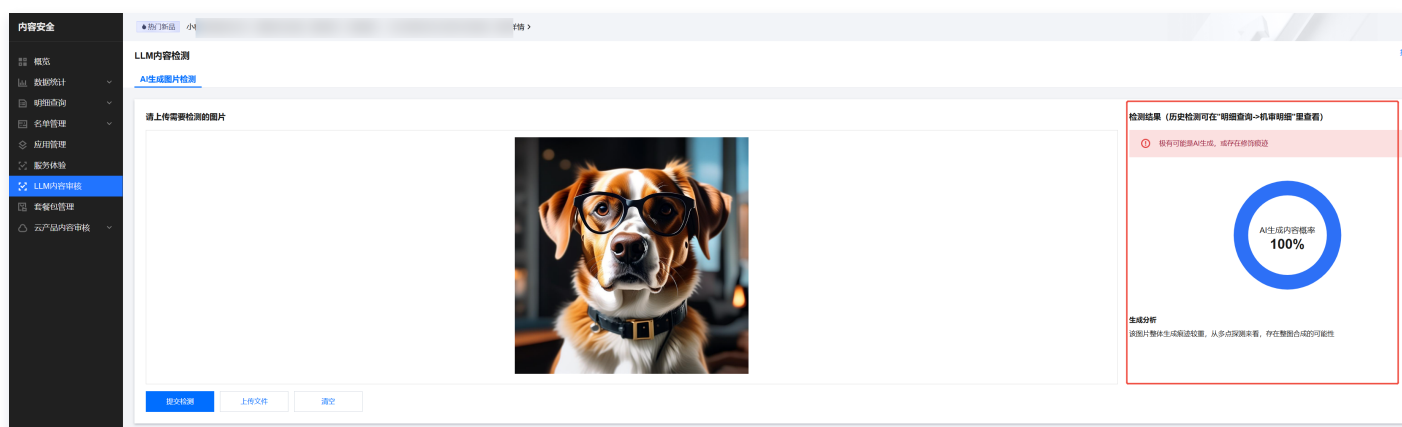
- 登录 [内容安全控制台](#)，单击 LLM 内容审核，跳转后单击初始化配置，立即体验。



2. 在指定区域单击上传文件或将文件拖拽到此区域



3. 上传文件或文件拖拽到此区域后，单击提交检测，即可查看检测结果。



方式二：通过 API 接口调用

若您需要和您的业务系统做对接，则可使用 API 接口的方式来调用我们的服务。

- 开通服务：**登录 [内容安全控制台](#)，单击 **LLM 内容审核**，跳转后单击**初始化配置**，立即体验。
- 获取接口参数：**单击“**应用管理**”，即可见如下图所示的“**AI 生成检测配套策略**”，其中 **Biztype** 字段对应的值，即接口调用的必备传值。
- 系统对接：**请参见 [图片内容安全接入文档](#) 来完成系统接口的对接，其中，接口参数中 Biztype 的传值，即为上一步您在“应用管理”页面查看的 Biztype 值。

