内容安全

产品简介

产品简介

什么是内容安全

背景

随着互联网、智能设备及各种新生业务的飞速发展，互联网上的数据呈现爆炸式增长，图片、视频、发文、聊天等互动内容已经成为人们表达感情、记录事件和日常工作不可或缺的部分。每天，通过互联网上传的视频、图片数量超过10亿，通过各种社交网络、媒体平台的发文数量超过5亿，而且这种趋势还是继续快速增长中。

这些日益增长的内容中也充斥着各种不可控的风险因素，比如色情视频和图片、涉政暴恐内容、各种垃圾广告等等，随着政府监管的日渐严格，这些都是各网站及平台亟待认真对待和管理的工作。而另一方面，人们对这些非结构化内容的认识和解读和也处于初级阶段，需要更加智能的技术和系统来帮助大家深度发掘这其中蕴藏的巨大商业价值。

产品定位

云盾内容安全是内容安全领域的先行者，源自阿里巴巴多年安全技术积累，依托阿里云、淘宝、支付宝等平台的管控经验，为企业用户提供成熟的、轻量化接入的内容安全解决方案，帮助企业、开发者在复杂多变的互联网环境下快速发现文本、图片、视频的各类风险，保障应用的信息内容安全。

目前的产品包括：内容检测API，OSS违规检测，ECS站点检测。

内容检测API

内容检测API主要对包含色情、涉政、暴恐、广告、垃圾信息的文本、图片及视频进行检测和识别，通过系统化的方式提供审核、打标、自定义配置等能力来保障您接入的效果和个性化需求落地。

针对的用户包括但不限于：视频网站、直播平台、社交平台、媒体平台、垂直社区 / 论坛、电商网站、存储平台、CDN平台等一切UGC（用户生成内容）平台和一切需要对本站内容进行安全管控的平台。

ECS站点检测

针对阿里云ECS用户，提供首页检测服务和网页内容检测服务，帮助您检查您的首页是否具有被攻击、挂马等风险，以及当您的站点中网页疑似有违规信息时，会通知您并提供违规网页地址及快照查看功能，方便您对网

页内容进行整改。

OSS违规检测

针对阿里云OSS用户，提供一键式的图像鉴黄SaaS化服务。您可以将保存在OSS中的图片进行鉴黄检测，并且提供删除和冻结的功能。

本地化部署方案

内容安全提供本地化部署版本。您可以将内容安全部署在本地，并对接本地数据中心，直接调用本地数据执行内容检测。本地化部署方案帮助您省却数据上传的工作，满足数据中心利旧和数据本地化需求。

内容安全本地化部署以软件方式提供，购买后由阿里云安全工程师到现场完成部署。通过本地化部署，您可以在自有数据中心（如自有IDC、物理机/私有云、混合云等）内获取阿里云云上环境同等量级的内容安全检测能力，也可以无缝获取公共云的弹性扩展能力，适应您的定制化要求和生态建设需要。

功能特性

内容安全本地化部署版本支持对**本地文本、图片、视频内容**进行特定场景的违规内容检测或特定内容识别，并提供管理控制台方便您进行相关配置和查看检测结果。

检测服务

以HTTP/HTTPS API接口形式提供指定场景的检测服务，便于您将检测环节集成到整体业务流程。支持的检测场景包括：

- **文本反垃圾**：采用NLP自然语言理解算法识别色情、暴恐涉政、广告、辱骂等文本垃圾，并且能够结合行为策略有效管控灌水、刷屏等恶意行为。
- **图片/视频色情识别**：对图片和视频进行色情内容识别以及色情程度量化。
- **图片/视频涉政暴恐识别**：识别暴恐旗帜、人物和场景以及敏感政治人物等风险信息。
- **图片/视频敏感人脸识别**：提供包括政治人物、敏感人物、以及名人明星等人物的面部识别，能够避免业务的违规和侵权风险。
- **OCR图文识别**：识别图片中的文字，精准定位图片中文字位置，准确识别斜排字、艺术字等字体。

管理控制台

管理控制台帮忙您执行以下操作：

- **内容安全私有化管理**，包括用户管理、系统配置（算法业务策略配置）、运维信息管理、调用报表查询等。

- **内容安全检测配置**，包括自定义图库、词库管理，检测结果查看、反馈等。

购买及部署方案

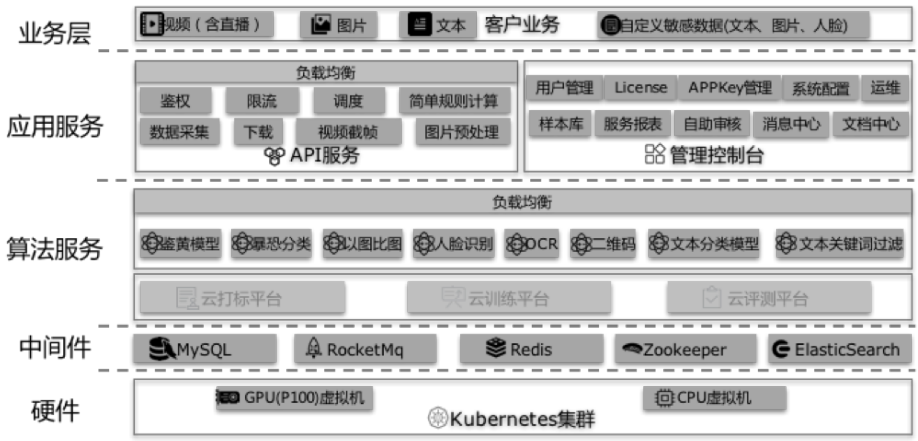
如何购买和收费

内容安全本地化部署版本目前开放**线下购买**途径，如有需求，您可以联系阿里云客户经理进行咨询和开通（您也可以提工单联系我们）。

本地化部署版本按照算法模型数量、服务量级及服务时间等收取相应的软件及服务授权费用，详情请咨询阿里云客户经理。

如何部署

下图阐释了内容安全的本地化部署架构。



在您购买内容安全本地化部署版本，并根据推荐的硬件配置要求准备好相应的硬件之后，阿里云安全工程师将到现场帮助您完成部署。

功能特性

内容安全主要包括以下功能：

- **站点监测服务**：内容安全针对阿里云网站类用户，提供信息内容安全检测及管控服务。当您的网站内容涉及违规信息时，会提前预警，并提供违规网页地址及快照查看功能，免去您手动检测网站内容烦恼，轻松解决网站违规信息。
- **OSS违规检测服务**：通过人工智能技术对用户存储在OSS上的图片进行违规识别，并对您提供便捷易用的结果展示平台。通过删除、忽略等快捷操作，方便您对图片作快速处理，减少审核人力，有效降低涉黄风险。
- **内容检测API**：内容检测API基于阿里巴巴多年的技术沉淀和海量的数据支撑，提供文本、图片、视频

等多媒体内容安全检测的开发接口服务。该服务可不依赖于阿里云其他服务，只要是公网可访问的图文信息均可过滤。

产品优势

- **性价比高**：在节省90%以上的人力成本的同时，支持秒级返回结果，达到99%以上的准确率。
- **经历实战检验**：支撑阿里系淘宝、支付宝等核心业务，经历“双11”实战检验，拥有海量的特征样本及丰富的数据模型分析经验。
- **接入成本低**：一次接入即可提供音视频、图片、文字等形式内容检测，覆盖暴恐、鉴黄、涉政、广告等风险防范。
- **灵活的服务方式**：既与OSS、ECS等云产品无缝对接，又可以通过API方式与用户审核系统集成。
- **海量数据快速检测**：基于云计算平台，对海量数据进行快速检测。

发展历史

- 2015年12月3日，内容安全内容检测API服务正式上线。
- 2015年10月22日 内容安全OSS违规检测服务正式上线。
- 2015年5月12日 内容安全ECS站点检测服务正式上线。

名词解释

站点检测

- **网站首页检测**：首页检测服务可以帮助您检查您的首页是否具有被攻击、挂马、无法访问等风险。
- **网页内容检测服务**：网页内容检测服务针对阿里云网站类用户提供网页检测服务。当您站点的中网页疑似有违规信息时会通知您，并提供违规网页地址及快照查看功能，方便您对网页内容进行整改。

OSS违规检测

- **违规分值**：代表趋近于违规图片的概率，分值越高，是违规图片的概率越大。
- **冻结**：冻结OSS object意味着该图片无法在外网被访问。