



2016 杭州·云栖大会  
THE COMPUTING CONFERENCE

云栖社区  
yq.aliyun.com

# 内容安全检测与管控

2016  
The Computing Conference

主办单位:



战略合作伙伴:



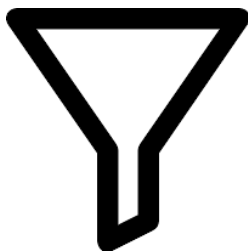
扫码观看大会视频

## 1 背景介绍



- 大背景
- 小行业

## 2 内容检测



- 简单高效
- 接入门槛低

## 3 困难与思路



- 客户又快又猛
- 标准多种多样

## 4 一些计划



- 一些想法
- 一些动作





Facebook每日新增  
600T数据



Twitter每日新增  
1.4亿条信息



Youtube每分钟新  
增300小时视频



## 1 缺乏内容安全管控经验

安全管控需要一整套的舆情情报、预警、分析、响应和沉淀机制

## 2 缺乏内容安全检测技术

研发成本高，需要不断应付变种违规内容，对抗逐步升级

## 3 安全管控耗费大量成本

研发、运营、审计，投入成本过高；尤其是早期全人审阶段





## 图片

经过多年积累、训练  
样本亿级别

针对多种场景特性，  
细分适配

检测能力：鉴黄、图  
文识别、人脸、二维  
码 ...

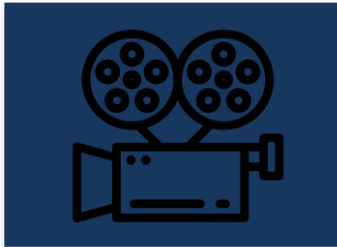


## 文本

词库丰富，基本覆盖  
各种场景

算法高效，违规内容  
实时识别

检测能力：敏感词、  
分词纠错、反垃圾 ...



## 视频

智能截帧，有效抓取  
嫁接违规视频

检测能力：鉴黄



## 网站

依靠阿里云多年实战  
经验，不断积累/沉  
淀管控策略

与阿里云产品联动，  
更高效、更灵活

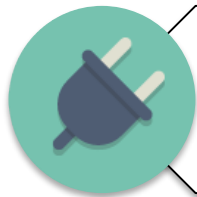
检测能力：挂马、网  
站可访问性 ...





## Open API

Http + 多语言SDK，更加通用



## Plug-in

逐步适配多种建站系统，有效降低接入成本



## SaaS

一键开通服务，提供简单方便的审核平台，所见即所得





快速接入



快速接入，满足80%需求

管控尺度不一



我要严管控，不能有漏网之鱼

行业特性不同



我是电商，商品黄图不一样

个性化服务



我是客户，我是上帝

通用模型

细化算分策略

提供不同模型

策略 + 模型







### 1 单一算法覆盖面有限

模型训练依赖样本集的收集、整理

### 2 模型升级与响应速度有待提升

新模型的训练与上线周期长

### 3 一个场景，多种算法支持

用户透明，效果更好，接入成本不变



## 1 能力边界

内容安全检测的能力边界清晰，可以与其它服务配合实现更好的效果

## 2 与CDN配合

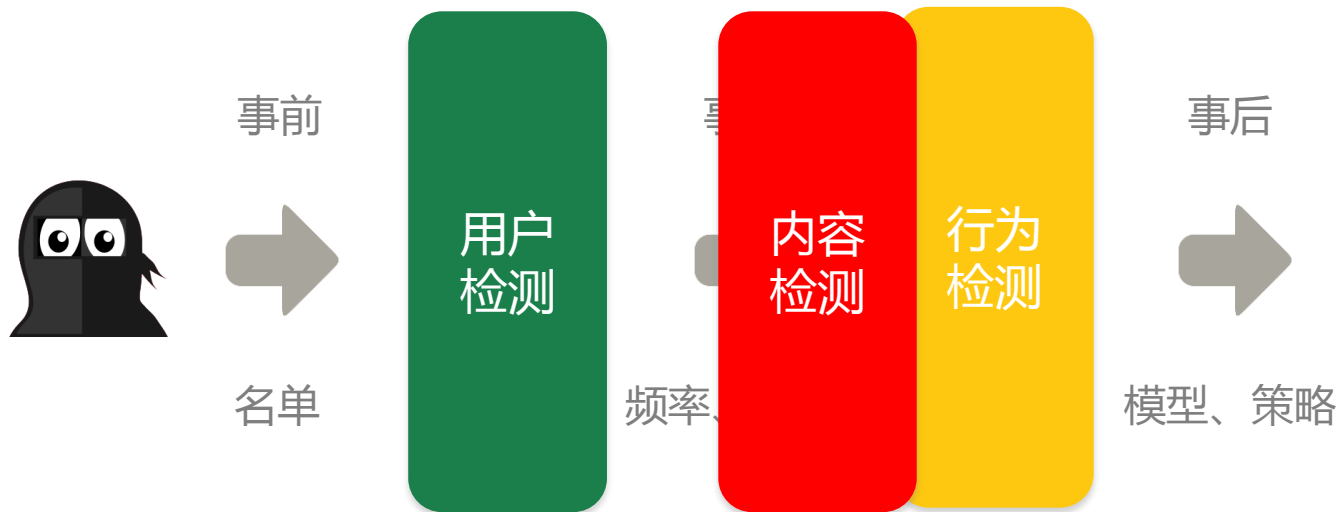
视频直播 + 鉴黄

## 3 与OSS配合

视频文件检测







2016 The  
Computing  
Conference  
**THANKS**

