文本内容安全

产品简介

产品文档

腾讯云

版权所有：腾讯云计算（北京）有限责任公司

【版权声明】

©2013-2024 腾讯云版权所有

本文档著作权归腾讯云单独所有，未经腾讯云事先书面许可，任何主体不得以任何形式复制、修改、抄袭、传播全部或部分本文档内容。

【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。

【服务声明】

本文档意在向客户介绍腾讯云全部或部分产品、服务的当时的整体概况，部分产品、服务的内容可能有所调整。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。

文档目录

产品简介
- 产品概述
- 产品功能
- 产品优势
- 应用场景

版权所有：腾讯云计算（北京）有限责任公司

产品简介

产品概述

最近更新时间：2023-12-20 16:02:45

内容安全审核面临的挑战

随着互联网、智能设备及各种新生业务的快速发展，互联网数据呈爆炸式增长，其中也充斥着各种不可控的风险因素，例如：色情内容、暴恐内容、垃圾广告等。
随着国家监管的力度不断提升，暴力、血腥、黄赌毒及危害青少年等不良内容被重点关注。内容安全审核成为短视频、新闻资讯、直播等平台急需的运营能力。
传统审核手段对于文本中常见的暴力性、讽刺性、暗示性的内容识别审核较难，审核成本较高，企业内容运营面临新的技术挑战。

什么是文本内容安全

腾讯文本内容安全（Text Moderation System，TMS）是一款文本内容智能识别服务，对用户上传的文本进行内容安全识别，能够做到识别准确率高、召回率高，多维度覆盖对内容审核的要求，并实时跟进监管要求，不停地更新审核服务的识别标准和能力。
能够对文本文件进行多样化场景检测，精准识别文本中出现可能令人反感、不安全或不适宜的内容，有效降低内容违规风险与有害信息审核成本。
能够精准识别涉黄、涉恐、暴力等有害内容，支持用户配置词库，打击自定义的违规文本。文本内容安全服务能检测内容的危险等级，对于高危部分直接过滤，对于可疑部分提交人工复审，从而节省审核人力，降低业务风险。
以开放API（Application Programming Interface，应用程序编程接口）的方式提供服务，用户通过调用API即可获取审核结果，高效构建智能化业务系统，提升业务运营效率。

版权所有：腾讯云计算（北京）有限责任公司

产品功能

最近更新时间：2023-12-20 16:02:45

色情内容识别

识别过滤多种涉黄违规类型的文本，包括低俗行为、隐晦涉黄、涉黄物体和性行为相关描述等内容。

暴恐内容识别

精准识别多种疑似暴力恐怖行为、物体（枪支、刀具、标志）和场景（暴乱、战争）等内容。

广告内容识别

过滤多种形式的广告文本，包括网络小广告、牛皮癣广告、招嫖广告等内容。

语种识别

当前仅支持中文。

自定义识别

支持用户自定义中文关键词库，精准识别包含自定义关键词的违规内容。您无须训练机器识别模型，系统会自动将识别内容与自定义库内样本进行匹配，对自定义内容进行定向识别，满足不同场景下的多样化审核需求。

版权所有：腾讯云计算（北京）有限责任公司

产品优势

最近更新时间：2023-12-20 16:02:45

高可靠性

业务可用性不低于99.9%，专业团队7 × 24小时实时提供技术支持。
请求毫秒级响应，结果秒级返回，超低延迟助力业务“快人一步”。
多集群部署，每秒超万级并发，支持动态扩容，无需担心性能损耗。

高灵活性

无需安装任何脚本文件，通过 API 即可直接调用业务接口，三步轻松接入。
支持自定义库和自定义审核策略，根据业务不同需求可以灵活配置内容安全服务。
支持腾讯云和非腾讯云客户使用，不涉及业务迁移成本。
支持**公有云服务**和**私有云部署**，全方位解决所有内容安全问题，省时、省钱、省心。

高性价比

采用后付费模式，按调用量扣费。
融合数十种算法技术构建综合识别模型体系，避免单易模型误判，机器确定部分准确率高达99.99%以上。

高可信任

腾讯22年产品运营经验与万级违规样本积累，模型学习样本丰富全面，覆盖各个行业数百种违规类型。

版权所有：腾讯云计算（北京）有限责任公司

应用场景

最近更新时间：2023-12-20 16:02:45

互动直播

腾讯文本内容安全可提供针对直播中的弹幕、用户评论等文本内容过滤的一站式、低时延解决方案，可对所有房间内容实时监控，识别可疑房间并进行预警。
如果已使用腾讯云的视频、直播解决方案，即可一键开启文本内容安全服务，及时阻断不良内容传播，降低平台运营风险。

社区论坛

腾讯文本内容安全可以广泛应用于博客、论坛等各类有用户原创内容的平台，包括个人主页、评论、发帖、回帖及站内信等场景。快速识别令人反感、不安全或有害内容，保障平台商业利益和业务合规，降低用户运营成本。

电商购物

腾讯文本内容安全可以对购物平台的文本内容进行全场景覆盖式识别，包括商品简介、商品详情介绍、买家评价和用户问答等场景。防止涉黄、涉暴、涉恐敏感类文本发布，降低人工审核成本和业务违规风险。
精细化的多级标签及高度客制化的自定义审核策略，保障各类合规商品正常交易的同时，及时甄别各场景下潜藏的不良和有害信息，维护平台业务的正常开展。

少儿教育

腾讯文本内容安全可以广泛部署在亲子生活、益智教育、在线教育、网络公开课等各类在线平台上，及时甄别线上教学、互动、录播课程中可能潜藏的各类有害信息。保障各年龄段用户尤其是未成年人的身心健康，营造良好的学习成长环境，提升用户体验。

即时通讯

腾讯文本内容安全服务依托技术领先的识别模型，精准识别消息中潜藏的各类可能令人反感、不安全或不适宜的内容，及其各种变体形式，并及时返回审核结果。
针对即时通讯的特殊场景能够有效防止恶意用户骚扰，阻断风险诈骗，提升用户体验。

版权所有：腾讯云计算（北京）有限责任公司