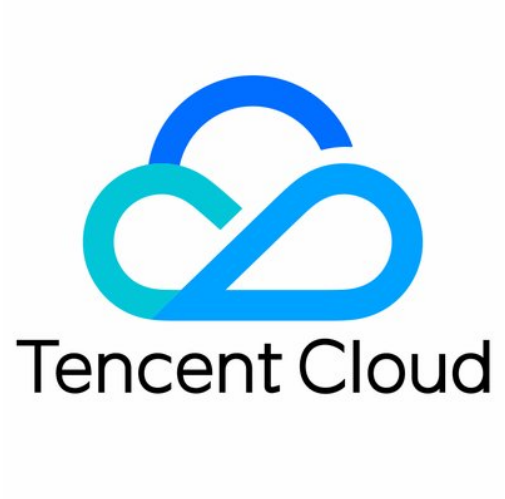
Text Moderation System

Product Introduction

Product Documentation



Contents

Product Introduction
Overview

Last updated：2023-12-20 16:02:45

Challenges in Content Moderation

With the rapid development of the internet, smart devices, and various emerging businesses, internet data has been growing explosively and is subject to diverse uncontrollable risks, such as pornographic text, terrorism content, and spam ads.
With ever heightened regulation, content involving violence, blood, porn, gambling, and drug use has become an area of focus, which makes content moderation an urgent need of short video, news, and live streaming platforms.
As common violent, ironic, and suggestive content in text are difficult and costly to recognize and moderate through traditional means, enterprises face new technical challenges in content operations.

TMS Overview

Tencent Cloud Text Moderation System (TMS) is a smart text recognition service for moderating the text uploaded by users. It delivers a high recognition accuracy and recall rate, meets the needs for content moderation in multiple dimensions, and has been constantly improved in its recognition standards and capabilities in response to the changing regulatory requirements.
It can detect various scenes in text and accurately recognize content that may be offensive, unsafe, or inappropriate, which effectively reduces the risks of non-compliant content and moderation costs.
It can accurately recognize content that involves porn, terrorism, and violence and allows you to configure dictionaries to filter custom non-compliant text. It can also determine the risk level of content. Specifically, it directly filters non-compliant content and submits suspicious content for human review, so as to reduce the labor costs and business risks.
It provides services in the form of open APIs. You can call them to get moderation results and efficiently create a smart business system to increase the efficiency of business operations.

Features

Last updated：2023-12-20 16:02:45

Porn Recognition

TMS recognizes and filters text involving multiple types of porn, including vulgarity, obscure porn, pornographic objects, and description of sexual behavior.

Terrorism Information Recognition

TMS recognizes various types of figures, objects (e.g., guns, knives, and signs), and scenes (e.g., riots and war) suspected of violence or terrorism.

Ad Recognition

TMS filters various types of ads, such as normal product ads, spam ads, prostitution ads.

Language recognition

Currently only Chinese is supported.

Custom Content Recognition

TMS allows you to customize Chinese and English keyword dictionaries. It can accurately recognize non-compliant content that contains custom keywords. You don't need to train a machine recognition model, as the system automatically matches the content to be recognized with the samples in the custom dictionaries for targeted content recognition, thus meeting diversified moderation needs in different scenarios.

Strengths

Last updated：2023-12-20 16:02:45

High Reliability

TMS delivers a service availability of at least 99.9%. Its professional team provides 24/7 technical support.
It responds to requests in milliseconds, returns results in seconds, and delivers an ultra low latency, which helps your business grow faster.
It supports multi-cluster deployment with dynamic scalability and can sustain over 10,000 concurrent requests per second, so you don't need to worry about performance loss.

High Flexibility

TMS can be connected to in three steps by directly calling service APIs, with no need to install any script files.
It allows you to customize libraries and moderation policies so as to flexibly use the TMS service based on your specific business needs.
It can be used by both Tencent Cloud and non-Tencent Cloud customers without incurring any business migration costs.
It supports **public cloud services** and **private cloud deployment** to address all content moderation problems, which saves time, money, and worry.

High Cost Performance

TMS is postpaid and billed by the call volume.
It provides a comprehensive recognition model system created by integrating dozens of algorithms and technologies, which avoid false negatives produced by a single model and deliver a recognition accuracy of above 99.99%.

High Confidence

TMS offers models created based on Tencent Cloud's 22 years of experience in product operations and tens of thousands of samples of non-compliant text, which cover hundreds of violation types in different industries.

Use Cases

Last updated：2023-12-20 16:02:45

Interactive Live Streaming

TMS provides a one-stop, low-latency solution that filters text content in live streaming scenarios, such as on-screen comments and user reviews. It can monitor the content in all live rooms in real time to identify suspicious rooms and trigger alarms.
If you use Tencent Cloud video or live streaming solutions, you can quickly enable TMS to prevent the dissemination of non-compliant content and reduce the risks to platform operations.

Communities and Forums

TMS is widely used for various platforms such as blogs and forums with original user-generated content, including personal homepages, comments, posts, replies, and messages. It can quickly recognize offensive, unsafe, and harmful content in order to guarantee the platforms' business interests and compliance and reduce their user operations costs.

Ecommerce

TMS can recognize all scenes in text content on shopping platforms, including product descriptions and details as well as customer reviews, questions, and answers. This prevents them from posting text involving porn, violence, terrorism, or sensitive content and reduces the manual moderation costs and business violation risks.
While guaranteeing the normal sales of compliant products, refined multi-level tags and highly customized moderation policies can promptly detect hidden non-compliant and harmful information in various scenarios to safeguard the normal businesses.

Education

TMS can be widely deployed on various online platforms such as parenting, kid education, online education, and online open class to promptly recognize harmful information that may be hidden in online lectures, interactions, and recorded courses. It guarantees the physical and mental health of users at all ages, particularly minors, creates favorable learning and growing environments, and improves the user experience.

Instant Messaging

Based on technologically advanced recognition models, TMS can accurately recognize various types of content that may be offensive, unsafe, or inappropriate hidden in messages and their diverse variants. Then, it can promptly return the moderation results.
For instant messaging scenarios, it effectively prevents harassment by malicious users, thus blocking the risks of scams and improving the user experience.