LiveData 文本审核控制台使用手册

1、数据统计

1.1 数据概览

您可以在此总览该项目一段时间内检测趋势变化、检测语种分布、检出类型分布情况。



此处展示当日 0 时至当前时刻的检测条数、敏感内容条数、广告内容条数、敏感内容占比、 广告内容占比以及各数据项相较于上一自然日的环比变化情况。



1.1.2 检测条数&敏感内容条数&广告内容条数趋势

以上三个模块以折线统计图形式展示您筛选时间范围内检测条数、敏感内容条数、广告内容 条数随时间的数据变化趋势。

您可通过移动鼠标到折线上某一个点的方式,查看对应时间点的具体数据。

1.1.3 语种分布



此处以饼图的形式展示您筛选时间范围内,检测文本的语种分布情况。

您可通过移动鼠标到饼图上某一扇型区域的方式,查看对应语种的具体检测条数。



1.1.4 垃圾类型分布

此处以条形统计图的形式展示您筛选时间范围内,检出敏感内容的类型分布情况。

您可通过移动鼠标到某一矩形的方式,查看对应类型的具体检出条数。

1.1.5 垃圾内容&发布垃圾内容用户 top10

数据统计 - 数据模式 - 数据模式 - 敏感内容 - 广告内容 - 用户分析 -	25%	∩≇	12.50% ⊕%	12.50% 演演 发布垃圾内容的用户	6.25%	6.35% ₩EMA
策略配置	敏感内容 广告内容			敏感内容 广告内	容	
词库管理 · · · · · · · · · · · · · · · · · · ·	敏感内容	数量		用户ID	发布敏感内容的数量	占比
	一号领导	2		66666666	6	50.00%
	习近平	2		0	1	0%
	书记	1		81000001	1	1.59%
	他妈的	1				
	推女郎	1				
	李文亮	1				
	ควย	1				
	enculer	1				
	อมควย	1				
	[交友低俗]	1				

(1) 垃圾内容 top10

- 敏感内容:此处展示您筛选时间范围内被检出的敏感内容命中次数最多的前 10 个敏感词 及对应数量。

-广告内容:此处展示您筛选时间范围内检出的广告内容中出现次数最多的前 10 个广告内 容及对应数量。

(2) 垃圾内容用户 top10

若该项目在调用文本审核接口时传入了用户 id, 那么此处将分别展示您筛选时间范围内发送敏感内容、广告内容数量最多的前 10 个用户的 id、发送数量以及占其总共发送内容的比例。

(3) "—>" 按钮

- 点击"垃圾内容 top10"右侧的"—>",将进入「敏感内容」页面;
- 点击 "发布垃圾内容用户 top10" 右侧的 "—>",将进入「用户分析」页面。

1.2 敏感内容

数据统计	. 3	添加过滤条件+ 请输入关键词搜索			近24小时	最近7天 曲	2022-04-10 - 2022-04-11 -
数据概览							
敏感内容		批量操作					⊥ 下载
广告内容		敏感内容	数量	操作			
用户分析		一号领导	2	设为白名单			
策略配置							
词库管理	-	原句	过滤后			UID	日期
服务配置		习近平	***			66666666	2022-04-10 21:01:24
		在中共总书记习近平管治下不允许党内有不同意见,包括参与悼念 武汉医生李文亮等言论	在中******管治下不允许党内 汉医生***等言论	有不同意见,包	回括参与**武	66666666	2022-04-10 21:01:16
		习近平	2	设为白名单			
		书记	1	设为白名单			
		他妈的	1	设为白名单			

该页面按照检出敏感内容命中的敏感词分类统计数量。您可以在此快速浏览筛选时间段内命中不同敏感词的数据。

1.2.1 过滤条件

数据统计	- 添加过滤条件 +	请输入关键词搜索]		近24小时 最近7天
维度 担	叟索筛选条件	未选择任何过滤条件			
语言	□ 阿拉伯语				
类别	 德语 英语 		数量	操作	
	□ 西班牙语		2	设为白名单	
	 法语 泰语 		2	设为白名单	
	中文(简体)				
	□ 中文 (繁体)		1	设为白名单	
			1	设为白名单	
		取消 确认	1	设为白名单	
	李文亮		1	设为白名单	
	ควย		1	设为白名单	

您可通过勾选语言、命中敏感类别, 输入关键词, 选择时间四个维度共同筛选要查看的敏感 内容。

在此处修改所选时间段,下方将展示您所选时间段对应的敏感数据。

1.2.2 设为白名单

数据统计	- 添加过滤条件 + 请输入关键词搜索		近24小时 最近7天 · · · · · · · · · · · · · · · · · · ·
策略配置 词库管理 ·	批量操作		不必
黑名单管理	敏感内容	数量	操作
白名单管理服务配置	一号领导	2	设为白名单
	习近平	2	设为白名单
	书记	C10	设为白名单
	他妈的	1	设为白名单
	推女郎	0 1 0	设为白名单
	李文亮	1	设为白名单
	ควย	ି 1 ି	设为白名单

您可通过此功能将敏感词设置为"白名单"词汇。

点击每个词汇后的"设为白名单"按钮,系统将提示是否确认执行操作。在您确认后,对应 敏感词将进入白名单,您可在「白名单管理」页面查看。同时,系统对仅命中该敏感词的文 本将不再检出。

1.2.3 批量操作

数据统计 •	添加过滤条件+ 请输入关键词搜索		近24小时 最近7天	11 -
数据概览 敏感内容	取消 设为白名单		ط ا	载
广告内容	□ 敏感内容	数量	操作	
用户分析 策略配置	□ 一号领导	2	设为白名单	
词库管理 •	□ 习近平	2	设为白名单	
服务配置	☑ 书记	1	设为白名单	
	□ 他妈的	1	设为白名单	
	□ 推女郎	-1	设为白名单	
	□ 李文亮	1	设为白名单	
	ิ ควย	-1	设为白名单	

此功能可将敏感词批量设置为"白名单"词汇。

点击"批量操作"按钮,页面将进入如上图所示状态,您可勾选多个想要设置为白名单的词 汇。勾选完毕,点击"设为白名单"按钮并确认操作后的处理逻辑与 1.2.3 一致,同时该页 面自动退出上图所示状态。

在上图所示状态下,点击"取消"按钮,将直接退出该状态。

数据统计 -	添加过滤条件 + 请输入关键词搜索		近24小时 最近7天 菌 2022-04-10 - 2022-04-11
数据概览			
敏感内容	取消 设为白名单		靠工 不
广告内容	敏感内容	数量	操作
用户分析		2	设为白老单
策略配置		2	
词库管理 🔹	□ 习近平	2	设为白名单
服务配置	☑ 书记	1	设为白名单
	□ 他妈的	1	设为白名单
	□ 推女郎	1	设为白名单
	□ 李文亮	1	设为白名单
		1	设为白名单

此功能可根据您的筛选结果将对应具体文本内容以.csv 格式的文件下载至本地。

1.3 广告内容

数据统计 🔹	添加过滤条件 +				近24小时 最近7	天
数据概览						
敏感内容	语言	广告内容	置信度	数量	用户	时间
广告内容	英语	i mean 30m silver = \$30, 115m iron = \$30	80	1	66666666	2022-04-11 10:25:57
用户分析	中文 (简体)	若二维码扫,	90	1	81000001	2022-04-10 22:13:04
策略配置	常语	gubank com	100	1	81000001	2022 04 10 22:12:04
词库管理 🔹	关语	qubapk.com	100		81000001	2022-04-10 22-13-04
服务配置	中文(简体)	资源商去工作室打开资源传送过来	100	1	66666666	2022-04-10 21:01:08

该页面展示筛选时间段内检出的广告内容数据。与「敏感内容」页面基本相同,支持通过勾 选语言、命中敏感类别,选择时间三个维度共同筛选要查看的广告内容。

1.4 用户分析

数据统计 🔹	请输	入用户ID搜索					近24小	时 最近7天 🗐 2022-04-10 - 2022-04-11
数据概览 敏感内容	я	引户UID	4	检测条数		违规内容条数		操作
广告内容	6	6666666	1	12		8		宣看违规内容 ~
用户分析	>	一级类型	二级类型		原句	违规内容	数量	日期
策略配置 词库管理 -	ľ	涉政	中国政治		在中共总书记习近平管治下不允 许党内有不同意见,包括参与悼 念武汉医生李文亮等言论	书记, 总书记, 习近平, 一号领导, 一号领导, 悼念, 李文亮	1	2022-04-10 21:01:16 "რუ
服务配置		每站	谩骂人身攻击		他妈的	他妈的, [交友低俗]	1	2022-04-10 22:57:59
		色情	其他色情		推女郎	推女郎	1	2022-04-10 22:57:22
		广告	工作室广告		i mean 30m silver = \$30, 115m iron = \$30		1	2022-04-11 10:25:57
		涉政	中国政治		习近平	习近平, 一号领导, 一号领导	1	2022-04-10 21:01:24
		违禁	毒品迷药		太白粉	白粉	1	2022-04-10 22:58:47
		广告	工作室广告		资源商去工作室打开资源传送过 来		1	2022-04-10 21:01:08
		帶路	谩骂人身攻击		vasi j'arrete ca m'énerve je pas se trop de temp à jouê juste po ur se faire enculé par des conn ards	connard, enculer	1	2022-04-11 10:25:39
	8	1000001	6	63		3		查看违规内容

默认情况下, 该页面以用户 id 为主维度展示对应用户在筛选时间范围内检测内容数量、违规内容数量以及具体文本内容。

^{1.4.1} 筛选条件

>
5
>

您可从输入用户 uid 和选择时间两个维度筛选要展示的内容。

1.4.2 展示字段选择

数据统计 -	请输入用户ID搜索			近24小时 最近7天 🗎 2022-04-10 - 2022-04-11	
数据概览 納威内容	用户UID	检测条数	违规内容条数	操作	
广告内容	展示字段:	12	8	宣看违规内容	>
用户分析	用户UID 检测条数	63	3	查看违规内容	>
策略配置 词库管理 -	违规类型 ■ < 违规内容条数 ✓				
服务配置					

您可通过上图所示的"展示字段配置器"选择要展示的字段。当选择展示"违规内容占比"时,系统将在"违规内容条数"列后增加展示"违规内容占比"列;当选择展示"违规类型"时,系统将根据"用户uid+类型"对筛选数据结果分行展示。

2、策略配置

用于针对同一个项目中的不同场景,配置审核松紧程度不同的策略。

2.1 默认策略

数据统计 ▼	创建策略				
词库管理	请输入策略名称搜索				
服务配置	策略编号	策略名称	应用场景	最近修改时间	操作
	03	简介	简介	2022-01-19 16:17:51	编辑 验证 删除
	02	社区	社区	2022-01-19 16:17:23	编辑 验证 删除
	01	昵称	昵称	2022-01-19 16:16:57	编辑 验证 删除
	DEFAULT	默认策略 ⑦	默认	2022-03-31 10:15:37	编辑 验证 删除

项目创建时,系统自动创建的策略。当调用文本审核接口的入参 strategyld 传入为空值时,将默认执行该策略。

注:默认策略不支持删除。

2.1.1 策略编辑

数据统计 -	← 返回策略列表				
策略配置	文本策略编辑				
词库管理 - 服务配置	 策略编号: DEFAULT 策略名称: 默认策略 应用场景: 群聯 	电接触数 以下公共可念译 德克小泽塔 一个			
	请依据忘时关除亚方而不近洋而安配亘时又4	ト甲依東朝,以下刀矢可多匹,而主之匹持一			
	🗾 涉政 ⑦	展开			
	☑ 暴恐 ⑦	展开			
	☑ 违禁 ⑦	展开			
	☑ 色情 ⑦	展开			
	☑ 辱骂 ⑦	收起			
	详细类别		语言配置		
	☑ 谩骂人身攻击		全部语种 🗸		
	☑ 不文明用语		全部语种 🗸		
	☑ 其他辱骂		全部语种		
	☑ 仇恨言论 ⑦	展开			
	☑ 未成年保护 ⑦	展开			
	☑ 敏感热点 ②	展开			

点击"编辑",进入「文本策略编辑」页面,可调整对应策略开启检测的类别。

2.1.2 策略验证

文本策略验证	×	
请输入文本	×	
	操作	
q	编辑 验证 删除	
c	编辑 验证 删除	
C	9000 编辑 验证 删除	
C	编辑 验证 删除	
开始检测		0
		6

点击"验证",可输入文本测试您配置策略的检测效果。

2.2 创建策略

数据统计 · ▼	创建策略				
词库管理	请输入策略名称搜索				
服务配置	策略编号	策略名称	应用场景	最近修改时间	操作
	03	简介	简介	2022-01-19 16:17:51	编辑 验证 删除
	02	社区	社区	2022-01-19 16:17:23	编辑验证删除
	01	昵称	昵称	2022-01-19 16:16:57	编辑 验证 删除
	DEFAULT	默认策略 ⑦	默认	2022-03-31 10:15:37	编辑验证删除

点击"创建策略",将进入「文本策略创建」页面。

2.2.1 策略编号

数据统计 -	← 返回策略列表			
策略配置	文本策略创建			
词库管理 ▼	• 筆銘伯見 • 通約) 御殿絶見 今日 新今月後 主体2000年2010			
服务配置	油品建立 , 电晶化道描刷器, 车台, 整本用度, 不适为1-1,4,97,3			
	*策略名称: 请输入策略名称,汉字、字母、下划线组成,支持2-100字符			
	*应用场景: 请选择应用场景 ~			
	1週28的实际业务需求选择需要配置的文本审核演纂,以下分类可多选,需至少选择一个			
	■ 涉政 ① 展开			
	■ 暴恐① 展开			
	☑ 选禁 ① 展开			
	■ 色情 ① 展开			
	■ 尋覧① 展开			
	■ 仇假言论 ①			

您需要为此策略定义一个唯一编号, 该编号将作为区分不同策略的标识, 因此其与已有策略的编号不可重复。在调用文本审核接口时, 可将编号作为入参 Strategyld 的值传入, 系统将 调用对应策略检测文本内容。

2.2.2 策略名称

数据统计 -	- 返回策略列表	ł				
策略配置	文本策略创建					
词库管理 服务配置	*策略编号:	袁璐编号: 请输入策略编号,学母、数字组成、支持20个字符以内				
	*策略名称:	请输入策略名称,汉字、字母、下划线组成,支持2-100字符				
	•应用场景:	请选择应用场景 🗸				
	请根据您的实际	业务需求选择需要配置的文本审核策略,以下分类可多选,需至少选择一个				
	🗾 涉政 ⑦	展开				
	☑ 暴恐 ⑦	展开				
	☑ 违禁 ⑦	展开				
	☑ 色情 ⑦	展开				
	☑ 辱骂 ⑦	展开				
	☑ 仇恨言论	©				

您需要在此输入待创建策略的名称,以便于创建后快速查询、区分不同策略。

2.2.3 应用场景

数据统计 🔹	← 返回策略列表
策略配置	文本策略创建
词库管理 👻	
服务配置	*策略编号: 请输入策略编号,字母、数字组成,支持20个字符以内
	*策略名称: 请输入策略名称,汉字、字母、下划线组成,支持2-100字符
	• 应用场景:
	请根据您的实际、 昵称
	
	2 暴恐 ⑦ ^{弹幕} 私聊
	☑ 违禁 ⑦ 群聯
	自定义添加场景 ☑ 色情 ⑦
炒捉统计	← 返回策略列表
英語記書	
泉中記旦 词 広 巻理 -	文本策略创建 添加自定义场景 × · · · · · · · · · · · · · · · · · ·
服务配置	* 策略编号 :
	* 策略名称: 请输入
	取消 确定 * 应用场景: 自定义:
	请假接巡的实际业务需求选择需要配置的文本审核策略,以下分类可多选,需至少选择一个
	■ 涉政 ⑦ 展开
	■ 暴恐 ⑦ 展开

您可直接在系统定义的场景中选择,或者自定义一个场景名称。

以上三项均填写完毕后,可继续根据需要配置该策略要检测的类别;点击"保存"后,策略即创建完成。后续在调用文本审核接口时可通过策略编号使用对应策略。

2.3 策略搜索

数据统计 ◆ 策略配置	创建策略				
词库管理 🔹 👻	请输入策略名称搜索		Q		
服务配置	策略编号	策略名称	应用场景	最近修改时间	操作
	03	简介	简介	2022-01-19 16:17:51	编辑 验证 删除
	02	社区	社区	2022-01-19 16:17:23	编辑验证删除
	01	昵称	昵称	2022-01-19 16:16:57	编辑 验证 删除
	DEFAULT	默认策略 ⑦	默认	2022–03–31 10:15:37	编辑 验证 删除

当列表中存在很多策略时,您可通过在此输入策略名称快速找到需要查询的策略。

3、词库管理

黑名单用于自定义需要屏蔽的特殊敏感词;

白名单用于自定义系统认为是敏感但业务上无需屏蔽的非敏感词以及自定义发送内容无需 过文本审核的用户的 id。

3.1 黑名单管理

数据统计 - · · · · · · · · · · · · · · · · · ·	黑名单管理 您可以自定义关键词进行屏蔽,用户自定义的关键词只支持精确匹配。				
19)年管理 業名単管理 白名単管理	<u>添加关键词</u> 上传关键词				
服务配置	敏感词 广告词 全部语	┿ ▼ 关键词	查询	批量操作 🕹 下载	
	语言	敏感词	创建时间		
	不限语种	from concerned	2022-03-31 11:22:21	1	
	不限语种		2022-03-31 10:33:54	Ŵ	
	不限语种	-	2022-01-07 15:23:56	۵ 🗖	
	不限语种	102	2021-12-13 16:10:42		
	不限语种	former	2022-03-29 10:43:34		
	不限语种	10.00	2022-02-28 11:10:44	1	

您可通过直接添加和批量上传两种方式向黑名单中添加敏感词及广告词,添加后的敏感词及 广告词,将分类在下方列表中展示。

3.1.1 添加关键词

数据统计	← 返回关键词列表
策略配置	选择垃圾类型
词库管理	 ● 敏感词 ○ 广告词
黑名单管理	支持模糊匹配:
白名单管理	支持空格匹配:
操作记录	国家
服务配置	不限国家
	语言
	不限语种
	· · · · · · · · · · · · · · · · · · ·
	全部策略
	(4日)
	个通过
	一级类型
	用户自定义
	二级类型
	请选择
	关键词
	每行一个关键词,多个关键词请换行。每个词最多支持128字符。

点击"添加关键词",进入词汇添加页面。

(1) 选择垃圾类型

分为普通敏感词和广告词两种类型(注:黑名单中自定义的广告词需在策略配置中开启"用 户自定义广告"类别后,方可正常检测。)

(2) 国家

您可在此选择待添加敏感词适用的国家。添加成功后,对应敏感词将只在接口传入的 country 字段值与配置国家 code 相同时生效。

(3) 语言

表示敏感词(或广告词)适用的语种;即只有当检测文本中出现该敏感词且其语种识别结果 在该敏感词(或广告词)选中的适用语种中时,该敏感词(或广告词)才会被检出。

(4) 策略编号

表示敏感词(或广告词)适用的策略;即该敏感词只有在系统使用文本调用配置的对应策略 且文本中包含该敏感词时候,才会被检出。

(5) 结果

表示当检测文本命中该敏感词,系统返回的结果,包括:不通过和疑似两种情况。

(6) 类型

表示当检测文本命中该敏感词时,系统返回的检出类型,一级类型包括:用户自定义、涉政、 暴恐、违禁、色情、辱骂、仇恨言论、未成年保护、敏感热点、个人信息保护、私人交易、 违规表情;二级类型即为这些一级类下的子类。默认情况下,一级类型系统自动选中"用户 自定义"类,此时二级类型不支持选择。

(7) 关键词

要自定义的黑名单词汇。

(8) 是否报警

如开启,则表示当检测文本中包含该自定义词汇时,文本审核接口将返回 warning 字段,且 值为 true。(注:仅垃圾类型选择"广告词"时,可配置该功能)

点击 "保存", 系统将自动保存添加词汇, 并返回黑名单列表页。

3.1.2 上传关键词

数据统计 👻	←返回关键词列表
策略配置	24.42.5-47.米田
词库管理 🔹	広律业成実堂 ● 鶴應词 ○ 广告词
黑名单管理	
白名单管理	
服务配置	上(ス) 生の
	支持上传UTF-8編码的*txt和*csv文件。点击下载示例
	简要说明:
	 * 云上曲率支持您添加竞品游戏的关键词,并能根据这些关键词准确识别拉人广告。 * 您可以直接添加关键词也可以通过文件上传,只接受.csv文件,单个文件上线为2M。 * 文件中每一行为要识别的广告词及对应的语言,其中语言使用 iso-639-1 或 BCP-47。▲下载示例文件
	提示:保存成功后,策略预计5分钟后生效。
	保存

点击"上传关键词",进入词汇上传页面。

(1) 选择垃圾类型

同上。

(2) 覆盖已经存在的关键词

如开启,则在将上传敏感词保存后,系统将用本次上传的敏感词将历史所有添加过的敏感词 覆盖,仅保留本次上传结果。

(3) 上传关键词

点击"浏览"按钮,选择要上传的文件(文件格式需与"下载示例"中的文件格式一致)

点击 "保存", 系统将自动保存上传词汇, 并返回黑名单列表页。

3.2 白名单管理

数据统计 - 策略配置 词库管理 - 黑名单管理	白名单管理 项目透加白名单后,其中的内容不被判断为敏感词。 添加关键词 添加白名单用户				
白名单管理 操作记录 服务配置	自名单词 自名单用户 全部语种 ◆ 关键词 章询 批量操作				
	语言	关键词	创建时间		
	不限语种	₽jd	2022-05-05 15:39:32	0	

您可通过直接添加关键词的方式向白名单中添加要忽略的词汇或用户 id, 添加后的词汇活 用户 id 将在下方的列表中展示。

3.2.1 添加关键词

点击"添加关键词",进入词汇添加页面。

(1) 语言

表示该白名单词汇适用的语种;即只有当检测文本中出现该词被系统判定为"敏感",且其语种识别结果在该词选中的适用语种中时,该词才会再次被修正为"正常"结果。

(2) 关键词

要自定义的白名单词汇。

点击 "保存", 系统将自动保存添加词汇, 并返回白名单列表页。

3.2.2 添加白名单用户

↔五上曲率	文本單核 Text-Demo 数据总览	☷ 控制台	8
数据统计 -	←返回列表		
策略配置	田 户 ID		
词库管理 -			
黑名单管理	נראמואישו ומהן אב, שו נמהן - ברפּר .		
白名单管理			
操作记录			
服务配置			
	保存		

点击 "添加白名单用户" 按钮, 进入用户 id 添加页面。您可按照每行一个用户 id 的形式, 同时添加多个白名单用户。保存后, 系统对这些用户发送的内容, 将均返回通过结果。

3.3 操作记录

数据统计 -	◎ 开始日期	- 结束日期			
策略配置	操作动作	请输入操作人			
词库管理					
黑名单管理	敏感词	请输入操作词汇			
白名单管理	童询				
操作记录	操作词汇	操作动作	词汇分类	操作时间	操作人
服务配置	环	删除	敏感词	2022-07-15 12:07:16	
	zhaojianjun	删除	敏感词	2022-07-15 12:07:13	
	noluanmana	mite	dade mit 210	2022-07-15 12:07:10	vanmana naiMiliwadata com

操作记录中记录了用户对黑、白名单的操作数据。您可通过操作时间、操作动作、操作人、 词汇分类、操作词汇等筛选条件进行记录检索。

4、服务配置

数据统计 策略配置	•	基本信息		2 48	辑项目信.	<u>®</u>
词库管理	-	项目编号(pid)				
服务配置		项目名称	Text-Demo			
		项目分类	游戏			
		服务请求地址				
		旧版请求地址	warmen war dan in a men ten in CO-manufactor and a			
		密钥 ⑦ (下述两个密钥都可以)				
		#1 ******		ø	8	<
		#2 ***********		ø	e >	<
		其它设置		ŧ	请调整限	81
		是否付费?	Not Free			5
		每秒调用次数	10,000		4	

此页面展示项目编号、名称、分类、请求地址、密钥等信息。同时,您可通过"编辑项目信 息"按钮,修改项目名称、分类以及描述信息。