

AI审核消息功能说明

版本号：v1.0

一、功能概览

功能信息	说明
功能名称	AI审核消息
功能介绍	基于AI的聊天内容自动审核功能
使用目的	通过AI自动审核提升内容安全管理效率
使用场景	1. 大规模聊天内容需要自动审核 2. 需要提高审核效率 3. 需要统一的审核标准
操作路径	管理后台：默认模板/风监合规 -> 严禁词管理 -> AI审核消息 助教端：聊天室审核快捷开关

二、功能详情

1. AI审核机制

防录屏功能

AI助手答疑

直播回放

回放设置

视频库

直播统计

直播间统计

场次统计

高级功能

电话连麦

直播数据大屏

实时字幕

严禁词管理

聊天室

观网页设置

气氛助手

时移与看点

多会场

直播列表 / 合规风控 / 频道严禁词 (ai审核消息 - 5665696)

频道严禁词AI审核消息

用于直播聊天室的AI审核消息功能

- 开启后，聊天室消息将通过AI进行审核
- 审核不通过时，观众发送的消息将无法被他人看到。
- 新建频道时，可复用默认模板中的AI审核规则，提高配置效率。
- 请定期检查和更新AI审核规则，确保审核效果持续有效。

AI审核设置

AI消息审核：

开启后，聊天室消息将按照规则通过AI进行审核

AI审核规则：

输入审核规则，多条规则请换行输入

规则说明 0/500

对特殊角色生效：

开启后，讲师、助教、嘉宾的发言也会经过AI审核过滤

保存

1.1 审核规则体系

1. 平台基础规则（自动生效）

- 政治敏感词：涉及政治人物、历史事件、政策法规的敏感表述
- 暴力血腥词：煽动暴力、描述血腥场景的词语
- 色情低俗词：与色情内容相关的明示或暗示性词语
- 歧视性词语：针对种族、民族、性别、宗教等的歧视性表述
- 侮辱谩骂词：人身攻击、侮辱性的词语
- 非法活动词：涉及赌博、毒品、诈骗等非法活动的词语

2. 自定义规则

- 支持补充配置审核规则
- 与平台基础规则同时生效

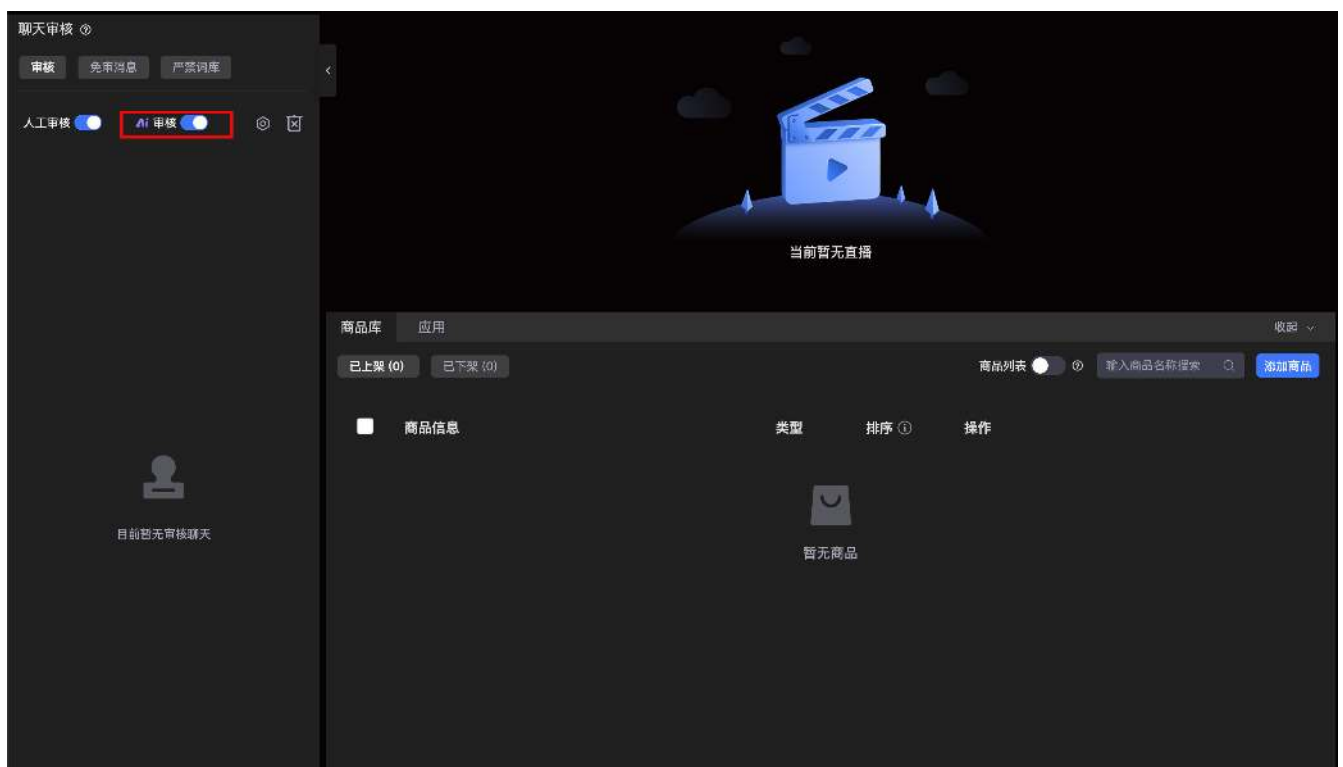
1.2 配置层级

1. 管理后台配置

- 平台模板配置（仅作为新建频道默认值）
- 频道独立配置（实际生效配置）

2. 助教快捷开关

- 位置：助教聊天室界面
- 支持快速切换AI审核状态
- 与管理后台配置同步生效



2. 使用说明

2.1 审核模式

- AI审核模式
 - 自动根据规则过滤内容
 - 适用于大规模聊天场景
 - 提高审核效率
 - 统一审核标准

注：系统支持切换至人工审核模式，两种模式可以共存

2.2 生效机制

- AI审核开启后立即生效
- 聊天内容自动经过规则过滤
- 违规内容自动屏蔽，其他用户不可见
- 支持实时切换审核状态

3. 功能效果

- 自动识别和过滤违规内容
- 提供统一的审核标准
- 支持自定义补充审核规则
- 实时生效，高效管理

三、注意事项

1. 规则优先级

- 平台基础规则不可关闭或修改
- 自定义规则作为补充，与基础规则同时生效
- 建议避免与平台基础规则重复

2. 配置说明

- 平台模板仅作为新建频道默认值
- 实际以频道配置为准
- 助教可通过快捷开关控制

3. 使用限制

- 与人工审核模式互斥,一条消息只能被一方审核
- 规则修改后立即生效

四、最佳实践

1. 使用建议

- 大规模直播场景优先使用
- 制定合理的自定义规则

- 定期评估规则有效性

2. 规则配置建议

- 结合业务场景设置规则
- 避免规则间冲突
- 保持规则的精确性

3. 日常维护建议

- 定期检查规则效果
- 及时更新补充规则
- 优化调整审核策略