

自定义音频审核规则填写指南

为了使 AI 音频审核模型能够准确识别您业务场景中的特定风险，建议您按照以下规范和技巧填写自定义审核规则。清晰、具体且逻辑严密的规则描述，能显著提升审核的准确率。

一、核心撰写原则

在撰写规则时，请遵循“场景+主体+具体行为”的逻辑，并明确界定“禁止”与“允许”的边界。

1. 明确具体行为，避免抽象描述

AI 难以理解模糊的形容词（如“禁止违规”、“禁止低俗”）。请将规则转化为具体的行为描述。

- 不推荐：** 禁止主播乱打广告。
- 推荐：** 在直播带货场景下，禁止主播口头引导观众添加个人微信、QQ 或加入第三方粉丝群进行私下交易。

2. 设定豁免条件（白名单）

当某些敏感词在特定语境下属于正常交流时，请务必说明“例外情况”，以减少误判。

- 推荐：** 禁止使用侮辱性词汇辱骂他人。**例外情况：** 游戏直播中，主播因游戏失利发出的感叹性语气词（未针对具体个人）不视为违规。

3. 提供典型话术样本

虽然 AI 具备语义理解能力，但提供行业典型的违规话术（样本）能帮助 AI 更快锁定目标。

- 推荐：** 禁止诱导未成年人刷礼物。**典型话术参考：** “小朋友把爸爸妈妈手机拿过来扫这个码”、“不用告诉家长”、“还没有上学的宝宝给我刷个礼物”。

二、分场景规则模版库

您可以根据实际业务类型，直接复制以下规则模版至输入框中，并根据自身需求进行修改或补充。

1. 电商带货与直播购物场景

管控重点： 违规引流、违反广告法、虚假承诺。

1. **禁止私下引流（导流）：** 严禁主播口播电话号码、微信号、QQ号，或通过“看主页简介”、“加粉丝群”、“领内部资料”等理由引导观众脱离平台进行私下交易。
2. **禁止绝对化用语：** 严禁使用违反广告法的词汇，如“国家级”、“最高级”、“第一”、“独家”、“绝无仅有”等词汇描述商品。
3. **禁止虚假承诺：** 严禁承诺“无效全额退款”（除非平台官方政策支持）、夸大产品功效（如减肥、医疗效果），或暗示售卖高仿、复刻产品。

2. 秀场颜值与社交交友场景

管控重点： 色情低俗、软色情ASMR、诱导非正规交易。

1. **禁止有声色情与ASMR：** 严禁主播发出呻吟、娇喘、舔耳、吮吸等具有强烈性暗示的声音，或进行ASMR表演。重点识别呼吸急促声及模拟性行为的声音。
2. **禁止传播色情诱导：** 严禁以“福利”、“私照”、“小视频”为诱饵，诱导用户刷礼物或加好友以换取色情或低俗内容。
3. **禁止低俗语言骚扰：** 严禁讨论性器官、性姿势、性经历，或对观众进行带有性意味的言语挑逗。

3. 游戏陪玩与电竞直播场景

管控重点： 辱骂攻击、违规代练交易、涉政言论。

1. **禁止恶意辱骂：** 严禁对队友、对手或观众进行人格侮辱、人身攻击或地域歧视。注意：游戏内的正常战术交流大声喊叫不属于违规。
2. **禁止违规交易宣传：** 严禁在语音中推销代练、卖分、外挂（脚本/科技）、低价代充等违规服务。
3. **禁止涉政涉恐言论：** 严禁讨论敏感政治人物、国家政策，或传播恐怖主义、极端主义言论。

4. 金融理财与股票证券场景

管控重点： 杀猪盘诈骗、非法荐股、虚假收益承诺。

1. **禁止承诺收益：** 严禁对金融产品做出“保本保息”、“稳赚不赔”、“年化收益率百分之百”等绝对化收益承诺。
2. **禁止非法荐股：** 严禁非认证人员自称老师或专家推荐具体股票代码，或引导用户加入“内部炒股群”、“VIP实盘群”。
3. **禁止诱导借贷：** 严禁推广无资质的借贷平台，或诱导学生、无收入人群进行贷款投资。

5. 情感咨询与连麦场景

管控重点： 不良价值观、封建迷信、精神控制。

1. **禁止传播不良价值观：** 严禁传播PUA教学（精神控制伴侣）、“杀猪盘”教学，或宣扬极端的拜金主义、物化异性等言论。
2. **禁止封建迷信活动：** 严禁以算命、看相、占卜、做法事为名义，诱导用户付费，或通过恐吓用户会有灾祸来骗取钱财。

6. 未成年人保护（通用高危规则）

管控重点： 诱导未成年人打赏、隐私侵害。

1. **禁止诱导未成年人消费：** 一旦识别到声音特征明显为未成年人，或用户自称为学生/未成年人，严禁主播索要礼物、现金或支付密码。
2. **禁止侵害未成年人权益：** 严禁对未成年人进行言语恐吓、诱骗、询问家庭住址及学校等隐私信息，或进行性骚扰。

三、进阶技巧：行业黑话与隐晦表达识别

若您的业务场景中存在大量“黑话”或“暗语”，请在规则中专门建立 **【黑话词典】** 模块，明确告诉 AI 这些词的实际含义。

填写示例：

...（此处填写常规规则）...

请特别注意识别以下行业黑话和隐晦表达，视为违规：

1. 代指金钱类：米、W、达不溜、软妹币、特产、棒棒糖（特定语境下代指资金）。
2. 代指联系方式类：绿色软件、V、企鹅、门牌号、神秘代码。
3. 代指违规品类：菠菜（博彩）、叶子（毒品）、气球（笑气）、上岸（网贷逾期处理）。

四、常见问题 (FAQ)

Q：规则写得越长越好吗？ A：不是。规则应追求精炼且准确。过多的冗余描述会干扰 AI 的判断逻辑，但必要的定义（如明确什么是“竞品”）不可省略。

Q：如何判断规则是否生效？ A：建议在规则保存后，关注后续的审核日志。如果发现存在漏判（违规未抓出）或误判（正常内容被抓），请根据实际案例微调规则描述，增加或减少限制条件。

Q：可以多条规则写在一起吗？ A：建议使用序号（1. 2. 3.）分条陈述，每一条规则对应一个独立的违规场景，这样有助于 AI 更清晰地理解逻辑关系。