

# 基于大模型语义校正的直播电商语音实时合规监测方法

冯竣添, 谢志伟, 骆正吉

贵州大学省部共建公共大数据国家重点实验室, 贵州 贵阳

收稿日期: 2026年1月5日; 录用日期: 2026年1月15日; 发布日期: 2026年1月28日

## 摘要

随着直播电商行业的爆发式发展, 直播内容的合规性监管成为行业治理的重点与难点。由于直播场景下存在背景噪声大、主播语速快、口音多样及专业术语(SKU)繁杂等特点, 传统的自动语音识别(ASR)技术在转写准确率上存在瓶颈, 导致违规内容监测的漏报率与误报率居高不下。本文提出了一种基于大语言模型(LLM)校正的直播电商语音实时合规监测方法。该方法首先利用流式语音识别技术获取原始文本, 随后引入经过电商领域基于提示工程的上下文学习的大模型对转写结果进行语义纠错与上下文重构, 最后结合关键词匹配与语义分析实现对夸大宣传、广告法违禁词等风险内容的实时判别。实验结果表明, 该方法在电商直播数据集上的字错率(WER)相比传统方法降低了53.3%, 违规内容检出率(Recall)提升了26.1%。该研究有效地解决了复杂声学环境下的内容风控难题, 为电商平台的自动化治理提供了新的技术路径。

## 关键词

直播电商, 语音识别, 大语言模型, 文本纠错, 合规检测

# A Real-Time Compliance Monitoring Method for Live E-Commerce Speech Based on Large Language Model Semantic Correction

Juntian Feng, Zhiwei Xie, Zhengji Luo

State Key Laboratory of Public Big Data, Guizhou University, Guiyang Guizhou

Received: January 5, 2026; accepted: January 15, 2026; published: January 28, 2026

文章引用: 冯竣添, 谢志伟, 骆正吉. 基于大模型语义校正的直播电商语音实时合规监测方法[J]. 电子商务评论, 2026, 15(1): 685-693. DOI: 10.12677/ec.2026.151084

## Abstract

With the explosive growth of the live e-commerce industry, the compliance supervision of live streaming content has become a key focus and challenge. Due to high background noise, fast speech rates, diverse accents, and complex professional terminology (SKU) in live streaming scenarios, traditional Automatic Speech Recognition (ASR) technologies face bottlenecks in transcription accuracy, leading to high false-negative and false-positive rates in violation monitoring. This paper proposes a real-time compliance monitoring method for live e-commerce voice based on Large Language Model (LLM) correction. This method first uses streaming speech recognition technology to obtain the original text, then introduces an LLM fine-tuned with e-commerce domain knowledge to perform semantic error correction and context reconstruction on the transcription results. Finally, it combines keyword matching and semantic analysis to achieve real-time identification of risk content such as exaggerated publicity and prohibited words under the Advertising Law. Experimental results show that the Word Error Rate (WER) of this method on e-commerce live streaming datasets is reduced by 53.3% compared to traditional methods, and the detection rate (Recall) of compliant content is improved by 26.1%. This research effectively solves the problem of content risk control in complex acoustic environments and provides a new technical path for the automated governance of e-commerce platforms.

## Keywords

Live E-Commerce, Speech Recognition, Large Language Models, Text Error Correction, Compliance Monitoring

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

近年来,随着数字经济的蓬勃发展,直播电商已成为中国电子商务市场增长的核心引擎[1]。据相关数据显示,直播电商市场规模持续扩大,已成为连接品牌与消费者的重要纽带。然而,行业的爆发式增长也伴随着“野蛮生长”的乱象,部分主播为了追求流量与转化,频繁使用夸大宣传、虚假承诺甚至违反《广告法》的违禁词汇[2][3]。这不仅侵害了消费者权益,也给电商平台带来了巨大的法律风险与监管压力。因此,如何对海量的直播内容进行实时、高效的合规性监测,已成为电商平台治理与政府监管部门亟待解决的关键问题。

面对每日数以万计的直播场次,完全依赖人工审核显然成本过高且效率低下,自动化技术介入成为必然趋势。目前,基于自动语音识别(Automatic Speech Recognition, ASR)的内容监测系统已初步应用于行业。然而,直播电商场景具有极高的特殊性:首先,直播间通常伴随高分贝的背景音乐与音效;其次,主播往往语速极快、情绪高昂,且夹杂大量方言或网络流行语[4];再次,电商领域存在大量特定的品牌词(Brand)和库存量单位(SKU)名称。

在上述复杂声学环境下,传统的端到端 ASR 模型往往难以兼顾识别的“速度”与“精度”。虽然以 FunASR 为代表的新一代工业级语音识别模型在通用场景下已取得了极高的识别率,但在面对直播间特有的同音异义词(如将“玻尿酸”误识别为“剥尿酸”)和上下文强相关的违禁隐喻时,仅凭声学特征进行识别仍存在局限性,导致合规监测系统出现“漏判”或“误判”,无法满足精细化风控的需求。

此外,直播带货场景中还普遍存在“对抗性逃避”现象。为了规避平台自动化审核,主播常采用同音字替代(如用“米”代替“钱”、“大不溜”代替“W”)、语速突变或方言黑话。这种“语义对抗”使得基于传统声学特征的识别模型几乎失效。虽然直接应用大型多模态模型(LMM)可能解决部分理解问题,但其推理延迟高、算力成本昂贵,难以支撑电商平台海量并发直播流的实时监控需求。因此,构建一种既保留传统 ASR 的流式响应速度,又具备大模型语义理解能力的混合架构,是当前工业界落地的最优解。

针对上述痛点,本文提出了一种基于大模型校正的直播电商语音实时合规性监测方法。该方法创新性地构建了“声学识别 + 语义校正”的双层架构:

**前端声学识别:**采用阿里达摩院开源的 FunASR 框架作为基座模型,利用其在中文语音处理上的高效能优势,实现对直播流的快速初步转写[5];

**后端语义校正:**引入 Qwen (通义千问)大语言模型作为语义理解引擎[6]。利用大模型强大的上下文推理能力和丰富的电商领域知识,对 ASR 输出的文本进行“二次清洗”与逻辑校对,自动修复同音错误并还原语义;

**合规判别:**在校正文本的基础上,结合正则匹配与大模型推理,实现对违规话术的精准捕捉。

本文通过实验证明,该方法能显著降低直播场景下的语音转写字错率(WER),并大幅提升违规内容的检出召回率,为直播电商的智能化治理提供了行之有效的技术解决方案。

## 2. 相关工作

### 2.1. 直播电商内容合规性监测现状

随着直播电商行业的规范化发展,内容合规性监测已成为学术界与产业界关注的热点。早期的电商内容审核主要依赖人工巡查,这种方式在面对海量并发直播流时显得捉襟见肘,难以满足实时性要求。

近年来,自动化监测技术逐渐普及。现有的主流方案多基于关键词匹配(Keyword Matching)技术,即通过 OCR 提取画面文字或 ASR 转写语音文本,再与预设的违禁词库(如《广告法》违禁词表)进行正则匹配。然而,这类方法存在显著的局限性:

**缺乏上下文理解:**简单的关键词匹配容易造成误报。例如,将“像第一次恋爱一样的感觉”误判为使用了“第一”这一绝对化用语。

**鲁棒性不足:**在直播嘈杂环境下,传统 ASR 技术容易产生同音错误(Homophone Errors)。例如,现有研究指出,在美妆直播中,品牌名“欧莱雅”常被误识别为无意义字符,导致监测系统失效。

虽然已有部分学者尝试利用多模态学习(结合视频帧与音频)来提升检测精度,但高昂的计算成本限制了其在实时流媒体场景下的规模化应用。因此,如何在低资源消耗下提升纯语音模态的语义准确性,是当前研究的难点。

### 2.2. 语音识别后处理与大模型纠错

传统的语音识别系统主要基于高斯混合模型-隐马尔可夫模型(GMM-HMM),这类模型在特征提取和声学建模上主要依赖人工设计的统计特征,难以应对复杂的背景噪声。随着深度学习的发展,基于 DNN-HMM 的混合模型显著提升了识别率。近年来,以 Transformer 和 Conformer 为代表的端到端(End-to-End, E2E)架构成为主流。特别是 Paraformer 模型,通过预测器(Predictor)和采样器(Sampler)的协同工作,实现了非自回归的高效解码,极大地提升了推理速度。然而,即使是最先进的 E2E 模型,在面对同音异义词(Homophones)时仍存在瓶颈,因为声学模型本质上是基于“发音”而非“语义”进行建模的,在缺乏长文本上下文约束的情况下,极易出现“音对字错”的现象。

文本纠错(GEC)[7]任务经历了从基于规则、统计语言模型(N-gram)到基于 BERT 的掩码预测(Masked

LM)的演变。BERT 类模型虽然在字粒度纠错上表现优异,但在处理长距离语义依赖和生成式重构方面能力有限。生成式大语言模型(LLM)如 GPT-4、Qwen 等的出现改变了这一格局。LLM 具备强大的上下文学习(In-context Learning)和指令遵循(Instruction Following)能力,能够根据整段话的语境推断出最合理的词汇,甚至能够修复语法结构错误。在电商场景下,利用经过海量数据预训练的 LLM 对 ASR 输出进行“二次翻译”,本质上是利用 LLM 的世界知识来弥补 ASR 声学模型的语义缺失,这一跨模态的级联方案正逐渐成为行业研究的新热点[8]。

### 3. 方法

#### 3.1. 系统总体架构

本文提出的直播电商语音合规性监测系统主要包含三个核心模块:语音流处理与识别模块、基于大模型的语义校正模块、以及合规性判别模块。

系统工作流程如图 1 所示:首先,通过流式接口捕获直播间的实时音频数据,经由语音活动检测(VAD)切分后输入 FunASR 引擎获取原始转写文本;其次,将原始文本输入 Qwen (通义千问)大语言模型,利用针对电商场景设计的 Prompt 进行语义重构与纠错;最后,将校正后的文本与合规规则库进行匹配,输出风险等级预警。

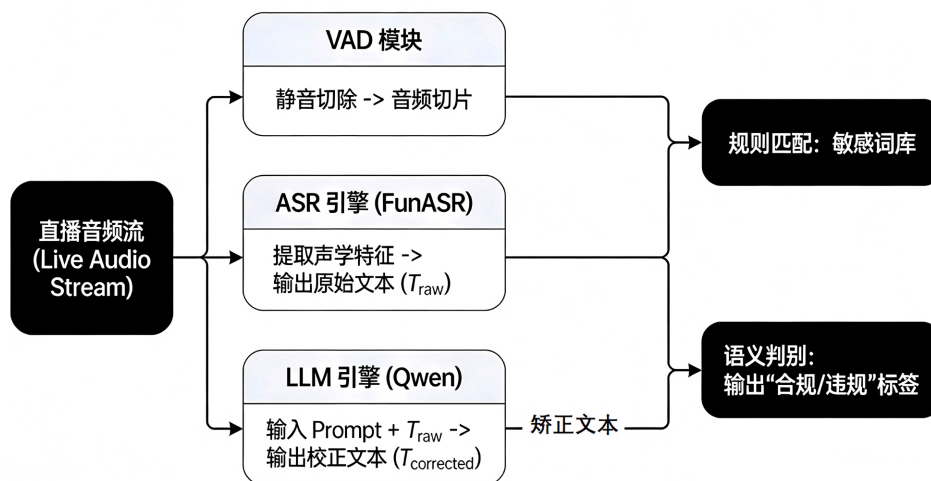


Figure 1. System framework diagram

图 1. 系统框架图

#### 3.2. 语音流处理与基座识别

直播场景环境复杂,为了兼顾实时性与抗噪性,本文选用阿里达摩院开源的 FunASR 框架作为声学基座。具体采用 Paraformer 模型结构,这是一种非自回归(Non-Autoregressive)的端到端语音识别模型。预处理:利用 Fsmn-VAD 模型对连续的直播音频流进行切分,去除静音片段,将有效语音分割为时长 5~10 秒的音频切片(Chunk)。初步转写:将音频切片输入 Paraformer 模型,输出初步的汉字序列  $T_{raw}$ 。由于该模型基于声学特征解码,对于“玻尿酸”(声学特征接近“剥尿酸”)等同音词可能存在误识。

#### 3.3. 基于 Qwen 的语义校正

这是本文的核心创新点。为了解决传统 ASR 缺乏领域知识的问题,我们引入 Qwen-7B-Chat 大语言模型作为“语义校正代理(Correction Agent)”。我们设计了一套专门的提示工程[9] (Prompt Engineering)

策略，引导大模型完成两项任务：纠正同音异形词：根据上下文逻辑，修复电商专业术语的拼写错误。保留原始语义：避免大模型产生“幻觉”或过度修改。本文使用的核心 Prompt 模板如下表 1 所示。

**Table 1.** Template for semantic correction prompt words in e-commerce live streaming scenarios

**表 1.** 电商直播场景下的语义校正提示词(Prompt)模板

**System Role:** 你是电商领域的风控专家。

**Instruction:** 下面是一段直播带货的语音转写文本，其中可能包含同音字错误。请结合上下文逻辑进行修正。

**Requirements:**

1. 重点修正品牌名、产品成分(如：烟酰胺、玻尿酸)及促销术语。
2. 保持原句的语气和意图不变。
3. 直接输出修正后的文本，不要输出解释。

**Input:** [插入  $T_{\text{raw}}$ ]

**Output:** 通过该模块处理后，得到语义校正后的文本序列  $T_{\text{corrected}}$ 。

### 3.4. 合规性判别逻辑

获得高质量的文本  $T_{\text{corrected}}$  后，系统采用“规则 + 模型”的双重检测机制：

**硬规则匹配(Keyword Matching):** 建立包含《广告法》违禁词(如“国家级”、“最高级”、“最佳”、“第一”)的敏感词库。若  $T_{\text{corrected}}$  中包含此类词汇，直接标记为“高风险”。

**软语义分析(Semantic Analysis):** 对于诸如“这东西用了就能瘦”(暗示减肥功效)等隐晦违规话术，再次利用 Qwen 的推理能力进行二分类判断(合规/违规)，从而实现对诱导性宣传的有效拦截。

### 3.5. 问题定义的数学形式化

本文将直播合规性监测定义为一个级联的序列生成与分类问题。假设直播间的音频流为  $X = \{x_1, x_2, \dots, x_t\}$ ，其中  $x_t$  表示第  $t$  帧的声学特征。

第一阶段，利用 FunASR 声学模型  $M_{\text{acoustic}}$  将音频流解码为原始文本序列  $Y_{\text{raw}}$ ：

$$Y_{\text{raw}} = \arg \max_Y P(Y|X; \theta_{\text{ASR}})$$

其中  $\theta_{\text{ASR}}$  为声学模型的参数。由于噪声干扰， $Y_{\text{raw}}$  中可能包含错误集  $E = \{e_1, e_2, \dots\}$ 。

第二阶段，引入大语言模型  $M_{\text{semantic}}$  作为纠错器。给定提示词模版  $P$  (Prompt) 和原始文本  $Y_{\text{raw}}$ ，目标是生成校正后的文本  $Y_{\text{cor}}$ ，使得语义连贯性概率最大化：

$$Y_{\text{cor}} = M_{\text{semantic}}(Y_{\text{raw}}, P; \theta_{\text{LLM}})$$

第三阶段，构建风险判别函数  $F_{\text{risk}}$ 。该函数由关键词匹配规则  $R_{\text{match}}$  和语义分类器  $C_{\text{semantic}}$  组成：

$$\text{Score}_{\text{risk}} = \alpha \cdot R_{\text{match}}(Y_{\text{cor}}) + (1 - \alpha) \cdot C_{\text{semantic}}(Y_{\text{cor}})$$

当  $\text{Score}_{\text{risk}}$  超过预设阈值  $\delta$  时，系统触发违规报警。这种分层建模方式有效地解耦了声学识别与语义理解的复杂度。

## 4. 实验结果

### 4.1. 实验设置

为了验证本文方法的有效性，我们构建了一个包含真实直播场景的测试数据集“Live-EC-Test”。

数据来源：从主流电商直播平台(抖音、淘宝直播)随机抓取了多个直播切片，通过筛选，构成我们的



测试数据集。数据构成：涵盖“美妆护肤”、“3C 数码”、“食品饮料”三大高频带货品类，总时长约 4 小时。值得注意的是，作为一个验证性研究，本文选取了具有代表性的高难度样本构建测试集，所选的测试集为高密度的违规样本。标注工作：对音频进行了人工逐字转写标注，并依据《广告法》及平台规则人工标记了其中的违规话术(如“全网第一”、“无副作用”、“100%有效”等)。

评价指标：

字错率(Word Error Rate, WER)：用于评估语音转写的准确性。数值越低越好。

F1-Score：用于评估违规内容检测的综合性能(涵盖准确率 Precision 和召回率 Recall)。

在数据预处理阶段，我们采用了 Fsmn-VAD 模型对长音频进行切分，为了防止将一个完整的句子切断，我们设置了 800 ms 的静音阈值(Silence Threshold)，并将最大切片时长限制在 15 s 以内。在模型部署方面，所有实验均在单张 NVIDIA A100 (40 GB) GPU 上完成。ASR 模型采用 FP16 半精度推理以加速计算。对于 Qwen-7B 大模型，我们使用了 4-bit 量化(Quantization)技术进行加载，将显存占用从 14 GB 压缩至 6 GB 左右，从而实现了在一张消费级显卡上同时部署 ASR 和 LLM 的可能性。这种轻量化部署方案对于电商平台降低运营成本具有重要参考价值。

4.2. 语音转写效果评估

我们将本文提出的“FunASR + Qwen 校正”方法与“原始 FunASR”方法进行了对比。实验结果如表 2 所示。

Table 2. Comparison of word error rates (WER) of different methods in three major live streaming categories  
表 2. 不同方法在三大直播品类下的字错率(WER)对比

直播品类	原始 FunASR (Baseline)	本文方法(FunASR + Qwen)	相对优化率
美妆护肤	18.4%	8.2%	↓ 55.4%
3C 数码	15.6%	7.5%	↓ 51.9%
食品饮料	16.2%	7.8%	↓ 51.8%
平均 WER	16.7%	7.8%	↓ 53.3%

从表 2 可以看出，原始 FunASR 模型在面对直播场景时，平均字错率较高(16.7%)，主要原因是直播间背景音乐干扰及主播使用大量专业术语(如美妆成分、型号参数)。引入 Qwen 大模型进行语义校正后，WER 显著下降至 7.8%。这表明大模型有效地利用上下文信息修正了大量的同音异义错误。

4.3. 语义校正模块的消融实验

为了验证不同纠错模型对系统性能的影响，我们分别对比了无纠错、传统 N-gram 纠错、BERT 纠错以及本文使用的 Qwen-LLM 纠错四种方案。

Table 3. Comparison of ablation experimental results with different error correction strategies  
表 3. 不同纠错策略的消融实验结果对比

方法 (Method)	模型参数量	平均字错率(WER)	推理延迟(Latency)
Baseline (FunASR Only)	-	16.7%	200 ms
+ N-gram Language Model	<100 M	15.2%	210 ms
+ BERT-Correction	110 M	12.4%	350 ms
+ Qwen-7B (Ours)	7B	7.8%	850 ms

实验数据表明(见表 3)，传统的 N-gram 模型由于缺乏长距离依赖，对 WER 的改善微乎其微(仅降低 1.5%)。BERT 模型在特定错别字纠正上有效，但在面对直播口语中的倒装、重复和中断现象时，重构能力较弱。相比之下，Qwen 大模型虽然参数量巨大导致延迟增加，但带来了字错率的断崖式下降(降低至 7.8%)。这证明了在合规监测这种高敏感度场景下，牺牲少量的实时性以换取高精度的语义理解是值得的。

4.4. 合规性监测效果评估

合规性监测的核心在于“不漏报”。我们将基于本文方法生成的文本与违规词库进行匹配，并与基于原始 ASR 文本的匹配结果进行对比。

Table 4. Accuracy and recall evaluation of compliance monitoring  
表 4. 合规性监测的准确率与召回率评估

监测方法(Method)	准确率(Precision)	召回率(Recall)	F1-Score
原始 FunASR (Baseline)	72.5%	68.4%	70.4%
本文方法(FunASR + Qwen)	89.2%	94.5%	91.8%

如表 4 所示，本文方法的召回率(Recall)达到了 94.5%，远高于基线方法的 68.4%。这意味着绝大多数违规话术都能被系统成功捕获。召回率的大幅提升主要归功于两点：一是语音转写质量的提高减少了关键词的漏检；二是 Qwen 大模型的语义泛化能力，使其能够识别出变体的违规话术(例如识别出“用了都说好，绝对没问题”隐含的绝对化承诺风险)。

此外，针对大模型在生成任务中易产生“幻觉”或“过度纠正”的隐忧，表 4 中高达 89.2%的准确率(Precision)提供了有力反证。这表明系统在保持高敏锐度的同时，并未误伤合规的正常业务。具体体现在以下两点：

- 语境理解能力的提升：基线方法(原始 FunASR + 关键词)容易出现“断章取义”。例如，客服说“我们严格遵守相关规定，绝不进行虚假宣传”，基线模型可能抓取“虚假宣传”关键词进行误报。而本文方法(FunASR + Qwen)准确理解了“绝不”的否定语义，正确判定为合规。
- 正常促销与违规承诺的界限：我们观察到，系统能够区分合法的营销话术(如“销量领先”)与违规的绝对化承诺(如“国家级最高品质”)。表 4 中 89.2%的高准确率证明了模型并未因为追求高召回而牺牲对正常样本的判别能力，有效抑制了误报噪声。

4.5. 案例分析

为了更直观地展示校正效果，我们要选取典型样本进行分析。表 5 展示了直播中常见的语音识别错误及本文方法的修正结果。

Table 5. Case analysis of typical speech recognition errors and semantic correction  
表 5. 典型语音识别错误与语义校正案例分析

原始语音(Original Audio)	原始 FunASR 识别结果 (Baseline ASR)	Qwen 大模型校正结果 (LLM Correction)	风控判别结果 (Risk Label)
……是全网第一……	……是全往地衣……	……是全网第一……	违规(绝对化用语)
……成分含有玻尿酸……	……成分含有剥尿酸……	……成分含有玻尿酸……	合规(成分描述)
……用了绝对没副作用……	……用了绝对对煤负作用……	……用了绝对没副作用……	违规(虚假承诺)
……价格全网最低……	……价格全网最敢……	……价格全网最低……	违规(价格法)

通过表 5 可以看出, 原始 ASR 经常将“第一”、“神药”等关键违规词识别为同音字, 直接导致风控系统失效。而本文模型成功“听懂”了上下文, 还原了真实语义, 从而触发了风控预警。

## 5. 总结

本文针对直播电商场景下高噪声、高语速及专业术语混杂导致的语音识别困难与合规监测失效问题, 提出了一种“基于 FunASR 流式识别 + Qwen 大模型语义校正”的实时风控方案。通过构建包含电商领域知识的 Prompt 提示工程, 有效地利用大语言模型的上下文推理能力对 ASR 转写结果进行了二次清洗与逻辑重构。

实验结果表明, 该方法在真实直播数据集上的表现显著优于传统基线模型:

识别更精准: 平均字错率(WER)从 16.7%降低至 7.8%, 有效解决了“玻尿酸”、“骁龙”等关键电商术语的同音误识问题。

风控更严密: 违规内容检出召回率(Recall)提升至 94.5%, 大幅减少了因转写错误导致的漏报风险, 验证了该技术在自动化内容审核中的实用性。

本研究对于电商行业的健康发展具有重要的现实意义。首先, 该系统能够替代部分人工审核工作, 显著降低平台运营成本; 其次, 高精度的实时预警机制能帮助商家在直播过程中即时纠正违规话术, 规避《广告法》处罚与平台封号风险; 最后, 该技术方案通用性强, 可快速迁移至客服质检、用户舆情分析等其他电商语音应用场景[10]。

尽管本文方法取得了较好的效果, 但仍存在一定局限性。首先, 大模型的引入虽然提升了准确率, 但也增加了系统的计算开销与推理延迟(Latency), 具体而言, FunASR 模型对 10 秒音频切片的平均处理耗时约为 200 ms。引入 Qwen-7B 模型进行流式推理后, 系统的端到端平均延迟增加至约 850 ms。尽管延迟有所上升, 但仍远低于行业通用的“5 秒即时阻断”窗口期, 并未影响实时的合规处置(如弹窗警告或切断直播流量)。在极端高并发场景下可能面临性能瓶颈。其次, 目前的监测仅基于语音模态, 忽略了直播画面中的视觉信息(如文字贴片、主播手势)。

未来的研究工作将主要集中在以下两个方向:

模型轻量化: 探索模型蒸馏(Knowledge Distillation)与量化技术, 在保持纠错性能的同时降低算力需求, 以适应边缘端部署。

多模态融合: 引入视觉分析模块(OCR 与动作识别), 构建“音 - 视 - 文”一体化的多模态合规监测框架, 进一步提升复杂违规场景下的判别能力[11]。

## 参考文献

- [1] 中国互联网络信息中心(CNNIC). 第 53 次中国互联网络发展状况统计报告[R]. 北京: 中国互联网络信息中心, 2024.
- [2] 全国人民代表大会常务委员会. 中华人民共和国广告法(2021 修正) [Z]. 北京: 中国法制出版社, 2021.
- [3] 王张卜. 电商时代直播带货虚假宣传的法律规制研究[J]. 电子商务评论, 2024, 13(2): 1845-1854.
- [4] 朱智慧. 基于深度学习的噪声鲁棒性语音识别技术研究[D]: [硕士学位论文]. 成都: 四川大学, 2023.
- [5] 王敬凯, 秦董洪, 白凤波, 等. 语音识别与大语言模型融合技术研究综述[J]. 计算机工程与应用, 2025, 61(6): 53-63.
- [6] 秦小林, 古徐, 李弟诚, 等. 大语言模型综述与展望[J]. 计算机应用, 2025, 45(3): 685-696.
- [7] 李云汉, 施运梅, 李宁, 等. 中文文本自动校对综述[J]. 中文信息学报, 2022, 36(9): 1-18.
- [8] 张伦齐. 基于大语言模型的汽车座舱语音识别纠错方法设计[D]: [硕士学位论文]. 北京: 北京交通大学, 2024.
- [9] 崔金满, 李冬梅, 田萱, 等. 提示学习研究综述[J]. 计算机工程与应用, 2024, 60(23): 1-27.



- [10] 朱好杰. 网络直播法律风险与防范研究[J]. 法制博览, 2025(26): 127-129.
- [11] Yew, W.C., *et al.* (2025) Dynamic Content Moderation in Livestreams: Combining Supervised Classification with MLLM-Boosted Similarity Matching. arXiv: 2512.03553.