SarcasmMiner：用于稳健视听讽刺推理的双轨后训练框架

Zhu Li[1]，*Yongjian Chen*[2]，*Huiyuan Lai*[2]，*Xiyuan Gao*[1]，*Shekhar Nayak*[1]，*Matt Coler*[1]

[1] 格罗宁根大学语音技术实验室，荷兰[2] 格罗宁根大学语言与认知中心，荷兰{zhu.li, yongjian.chen, h.lai, xiyuan.gao, s.nayak, m.coler}@rug.nl

arXiv:2603.05275v1 [cs.MM] 5 Mar 2026

摘要

多模态讽刺检测需要通过跨模态推理，解决文本、声学和视觉线索之间的语用不一致问题。为了借助基础模型实现稳健的讽刺推理，我们提出了 SarcasmMiner，这是一种基于强化学习的后训练框架，旨在缓解多模态推理中的幻觉问题。我们将讽刺检测重新表述为结构化推理，并采用双轨蒸馏策略：高质量教师轨迹用于初始化学生模型，而完整的轨迹集合则用于训练生成式奖励模型（GenRM），以评估推理质量。学生模型通过组相对策略优化（GRPO）进行优化，同时使用针对准确率和推理质量的解耦奖励。在 MUStARD++上，SarcasmMiner 将 F1 从 59.83%（零样本）、68.23%（监督微调）提升至 70.22%。这些结果表明，具备推理感知能力的奖励建模能够同时提升性能和多模态落地能力。

索引词：讽刺推理、后训练、强化学习、多模态大语言模型

1. 引言

大规模语音和多模态基础模型近年来在语音识别、语音合成和跨模态理解方面展现出令人印象深刻的能力 [1, 2, 3]。然而，仅靠预训练并不能保证模型在对话场景中对复杂语用现象进行可靠推理。因此，监督微调（SFT）和强化学习（RL）等后训练技术对于使基础模型适应需要结构化、基于依据的推理的下游任务而言，已变得至关重要 [4, 5, 6]。多模态讽刺检测是检验此类后训练方法的一项特别具有挑战性的任务。不同于传统的情感识别，讽刺源于字面意义与韵律、面部表情及上下文意图等非言语信号之间的不一致。例如，以夸张或平淡的韵律表达积极的词汇内容，可能暗示讽刺意图[7]。捕捉此类线索需要结构化的多模态推理，通常通过思维链（CoT）提示来实现 [8, 9]。然而，仅进行朴素的 CoT 监督并不能阻止模型捏造多模态证据 [10]，这凸显了推理感知强化学习的必要性。

在MUStARD和MUStARD++ 等数据集上，早期的多模态讽刺检测系统依赖于任务特定融合架构和监督分类目标[11, 12, 13, 14]。尽管这些方法行之有效，但它们是在多模态大语言模型（多模态大语言模型）出现之前发展起来的，因此并未探讨如何利用此类基础模型进行跨模态推理。近期工作已开始将推理监督纳入讽刺检测。Sarcasm-R1通过聚焦的推理策略增强文本讽刺检测，表明结构化推理能够提升可解释性与鲁棒性 [15]。然而，其表述主要以文本为中心，并未明确处理多模态落地。最近，MUStReason提出了一个用于评估视频大语言模型中语用推理的诊断基准 [16]。尽管它对感知—推理差距提供了有用的分析，但并未探索基于强化学习的后训练，也未探讨缓解幻觉式多模态证据的机制。

语音情感理解领域的并行研究已经开始探索以推理为中心的强化学习策略。Emotion-Thinker 将语音情感识别重新定义为一项感知韵律的推理任务，并引入韵律感知推理和带有渐进信任感知奖励的 GRPO，以提升可解释性和推理质量 [17]。GRPO引导的模态选择 [18] 和 B-GRPO [19]等相关方法进一步展示了策略优化在自适应模态使用和无监督样本选择方面的潜力。这些研究表明，强化学习能够挖掘多模态大语言模型超越简单标签预测的深层推理能力。然而，这些方法主要针对情感识别而设计，侧重于提升预测准确率或模态效率，而不是控制语用任务中的跨模态幻觉推理。尽管这些方法前景可期，但将其扩展到讽刺检测并非易事。首先，讽刺依赖的是细微的跨模态不一致，而非一致的情感线索。其次，多模态大语言模型可能会生成虚构的声学或视觉证据，为正确预测进行辩护。第三，不同于 EmotionCoT[17], 等情感推理数据集，大规模多模态讽刺思维链资源仍然不可获得，从而限制了直接的推理监督。

为应对这些挑战，我们提出 SarcasmMiner——一个基于强化学习、数据高效的后训练框架，旨在赋予全模态大语言模型具备抗幻觉的多模态推理能力，以用于讽刺检测。我们的框架不将讽刺检测视为纯粹的分类任务，而是明确对结构化的跨模态推理进行建模，并引入能够抑制幻觉式多模态证据的奖励机制。我们的贡献主要有三点：1）我们将多模态讽刺检测形式化为一个推理问题，并采用双轨蒸馏策略：正确的教师轨迹用于初始化学生模型，而包含幻觉的轨迹则用于训练奖励模型。2）我们提出一种生成式奖励建模范式，明确评估推理有效性，并惩罚虚构的音频或视觉证据，从而在预测准确率之外提升推理可靠性。3）通过全面的消融实验，我们证明，采用解耦的结果奖励与推理奖励的 GRPO 能够显著改善多模态落地，实现了高性能

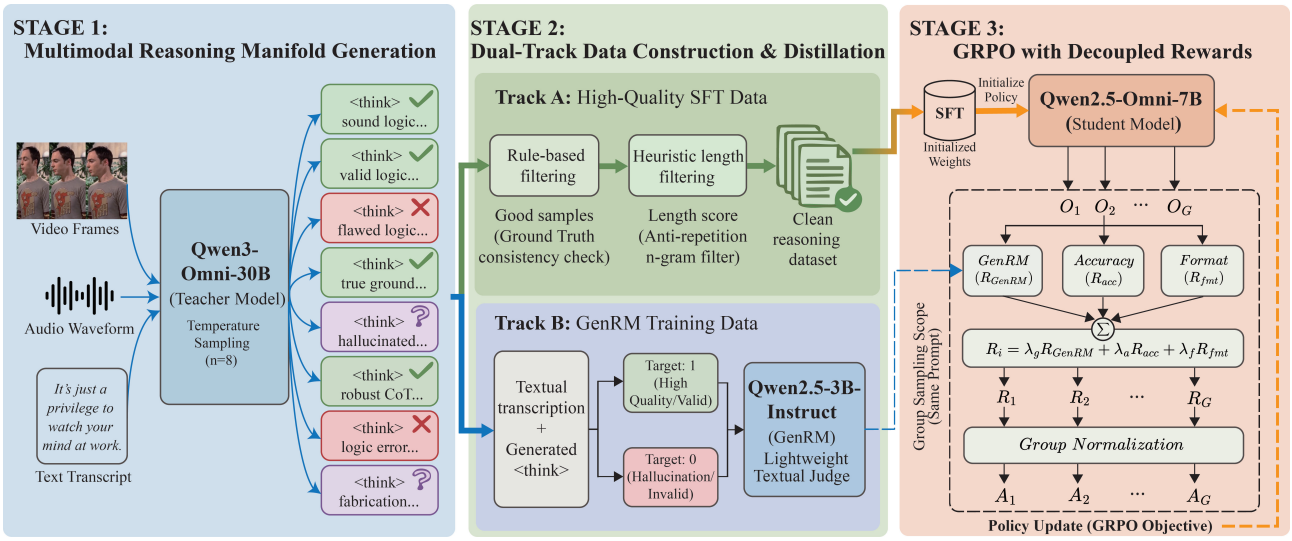

图 1： 所提出的 *SarcasmMiner* 框架概览 架构。

，在 MUStARD++ 上显著提高推理接受率。总体而言，本研究将基于强化学习的推理拓展至高层次语用推断，并为多模态基础模型的可信后训练提供了一条路径。

2. 方法

我们提出 SarcasmMiner，这是一种用于多模态大语言模型讽刺推理的三阶段后训练框架（图 1）。阶段 1 通过随机教师采样构建多样化的多模态 CoT 轨迹。阶段 2 执行双轨蒸馏：使用有效轨迹通过监督微调初始化学生模型，并使用完整的轨迹集训练生成式奖励模型（GenRM）。阶段 3 使用针对准确率、格式准确率和推理有效性的解耦奖励，通过 GRPO 对齐学生模型。

2.1. 阶段 1：多模态推理流形生成

多模态讽刺数据集通常缺少多步推理标注。为获得推理监督，我们提示强大的教师模型 $\mathcal{M}_{\text{teacher}}$ (Qwen3-Omni-30B)，分析多模态元组 $x = (v,\ a,\ t)$ 中转录文本与副语言线索（例如韵律和面部表情）之间的不一致，其中 v、a 和 t 分别表示视频帧、音频信号和文本转录。输出被限制为结构化格式：在 <think> 标签内包含推理链，随后在 <answer> 中给出最终的二元预测。我们不通过贪心解码生成单条确定性轨迹，而是针对每个输入 x，从教师模型的条件分布中采样，构建多样化的推理路径集合。具体而言，我们使用高温采样（温度 $T = 0.6$，top p $= 0.95$）独立采样 $n = 8$ 条轨迹，记为 $y_i \sim P_{\mathcal{M}_{\text{teacher}}}(\cdot|x)$，从而形成轨迹集 $\mathcal{Y} = \{y_1, \ldots, y_n\}$。这一随机过程生成了丰富的推理集合，其中包含正确推导、错误预测和幻觉式多模态推断，为后续双轨蒸馏提供了充足基础。

2.2. 阶段 2：双赛道数据构建与蒸馏 on

传统的基于拒绝的微调会舍弃次优轨迹，从而浪费其中蕴含的宝贵监督信号——这些信号来自失败案例。为解决这一局限，我们提出双轨蒸馏策略，系统地复用第 1 阶段生成的成功与有缺陷的推理路径。

赛道 A：高质量监督微调数据蒸馏。 为建立稳健的学生模型初始化，我们构建了一个“黄金”子集 $\mathcal{D}_{SFT}$。只有严格满足以下两项启发式标准的候选轨迹 $y_i \in \mathcal{Y}$ 才会被保留：1）与真实标签一致，即解析出的 <答案>与真实标签完全匹配；2）反重复，依据基于 n 元语法熵的筛选策略，过滤词元重复过多或生成熵过低的轨迹。对于每个输入，我们采用不同的轨迹选择策略构建 $\mathcal{D}_{SFT}$。具体而言，我们比较了三种初始化变体：1）贪心解码；2）best-of-N 采样（从 N 个候选轨迹中选出得分最高的轨迹）；3）多样化采样，保留多条正确且多样的轨迹[1]。

Track B：生成式奖励模型训练。 我们基于完整轨迹集 $\mathcal{D}_{Judge}$ 构建二元评判数据集，其中轨迹按质量标注，并训练生成式奖励模型，以区分高质量与低质量推理。每个回答 y_i 都会被赋予一个二元批判标签 $c_i \in \{0, 1\}$。我们仅将通过逻辑连贯的步骤得出正确标签的样本标记为 $c_i = 1$。$c_i = 0$ 则包括错误预测和多模态幻觉（例如，猜测正确但依据的是虚构的副语言证据）。我们通过监督微调，在 $\mathcal{D}_{Judge}$ 上微调轻量级语言模型 Qwen2.5-3B-Instruct，以获得 GenRM。与通过不稳定的价值头回归连续分数的标量奖励模型不同，我们的 GenRM 会评估推理上下文，并自回归地预测一个二元词元（“1”或“0”），从而为下游对齐提供稳定的监督信号。

2.3. 阶段 3：采用解耦奖励的 GRPO

基于赛道 A 得出的监督微调初始化，我们使用 GRPO 对学生模型 Qwen2.5-Omni-7B 进行对齐。GRPO 优雅地绕过了

[1]我们按比例增加贪心和 best-of-N 条件下的训练轮数，从而使所有模型的参数更新总次数相同。

传统强化学习评论家网络的内存瓶颈，这在处理内存密集型视听帧时是一项关键的架构优势。然而，仅针对标量准确率进行优化，会使策略极易受到奖励操纵的影响。具体而言，在多模态讽刺推理中，模型往往会利用统计捷径，捏造并不存在的声学或面部线索，仅仅为了证明其正确的标签猜测是合理的。为了明确惩罚此类多模态幻觉并确保逻辑有据可依，我们引入了一种解耦式奖励机制。对于多模态查询 x，策略 π_θ 会采样 $G = 8$ 条推理轨迹 $\mathcal{O} = \{o_1, \ldots, o_G\}$。每条轨迹 o_i 的总奖励定义如下：

$$R(o_i) = \lambda_a R_{acc}(o_i) + \lambda_f R_{fmt}(o_i) + \lambda_g R_{GenRM}(o_i) \quad (1)$$

具体而言，目标准确率奖励定义如下：

$$R_{acc}(o_i) = \mathbb{I}[\hat{y}_i = y], \quad (2)$$

其中，$\hat{y}_i$ 是从轨迹 o_i中提取的预测标签，而y 是真实标签。格式奖励会对格式错误的输出进行惩罚：

$$R_{fmt}(o_i) = \begin{cases} 1, & o \text{ satisfies the required format,} \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

最后，生成式推理奖励用于评估推理链的逻辑有效性。我们使用 Track B 中的 GenRM，将此奖励表示为生成正向验证令牌的自回归概率：

$$R_{GenRM}(o_i) = P_{\mathcal{M}_{\text{GenRM}}}(\text{token} = \text{“1”} \mid o_i). \quad (4)$$

将解耦奖励汇总后，GRPO 目标会计算组相对优势，以指导策略优化。这种归一化方式促使策略相较于同组的其他候选项不断改进。因此，能够正确预测且推理连贯、无幻觉的轨迹会获得更高的优势，并在优化过程中得到优先强化。这种组相对目标降低了对表面捷径的依赖，并促进模型依据可靠的多模态线索进行推理。

3. 实验设置

数据集。我们在 MUStARD++ [12] 上评估所提出的 SarcasmMiner 框架。MUStARD 是一个英语多模态讽刺检测数据集，包含 1,202 条带标签话语，涵盖文本、语音和视频模态，且讽刺与非讽刺样本的分布均衡。为确保不同数据划分中的类别平衡，我们按照 70/15/15 的比例随机将数据集划分为训练集、验证集和测试集，并使每个划分中的类别比例保持为 1:1。

评估指标。 我们从两个互补的角度评估模型性能：结果准确率和推理质量。对于结果准确率，我们衡量 <answer> 标签中的最终预测是否与真实值一致，并报告标准分类指标，包括准确率（Acc）和宏平均F1分数（Macro-F1）。为评估推理质量，我们衡量中间 CoT 轨迹的逻辑合理性。由于大规模人工标注推理路径并不实际，我们引入 **GenRM接受率（GAR）**作为一项自动指标。我们将微调后的 GenRM 用作测试集上的零样本评估器，并将 GAR 定义为被 GenRM 判定为有效（即预测 token“1”）的生成轨迹（来自 <think> 标签）的比例。GAR 是一种自动化的代理指标，用于衡量模型推理的逻辑一致性和多模态落地能力。

基线。 我们将所提出的框架与一系列在零样本设置下评估的先进全模态大语言模型进行比较。这些基础模型包括 Phi-4-Multimodal [20], 、MiniCPM-o-2.6 [21], 以及 Qwen-Omni 系列（Qwen2.5-Omni-7B Base [22], 、Qwen3-Omni-30B Instruct 和 Qwen3-Omni-30B Thinking [23]）。为系统评估所提出的后训练架构，我们将完整的 SarcasmMiner 模型与其消融变体进行比较，包括基于监督微调的变体（贪心、Best-of-N、多样化采样）和基于 GRPO 的变体（GRPO（无监督微调）、监督微调+GRPO）。

实现细节。 训练流程使用 ms-swift 框架 [24]实现，并通过 DeepSpeed ZeRO-3 进行优化 [25]。在监督微调阶段，学生模型和 GenRM 均使用 LoRA（$r = 8$、$\alpha = 32$） [26] 训练 4 轮，采用 AdamW 优化器，学习率为 2×10^{-5}。通过 GRPO 进行的强化学习后训练也运行 2 轮。在 GRPO 过程中，我们使用 vLLM 引擎为每个查询采样 $G = 8$ 条候选轨迹。策略以 1×10^{-5} 的学习率更新，并采用 $\beta = 0.01$ 的 KL 正则化系数，以约束其偏离监督微调参考策略。最终奖励是任务准确率、格式准确率和基于 GenRM 的推理质量的加权和，权重分别为 $\lambda_a = 1.0$、 $\lambda_f = 0.5$ 和 $\lambda_g = 1.0$。所有实验均在配备四块 NVIDIA H100 90GB GPU 的单节点上进行。

4. 结果与分析

表1报告了在 MUStARD++ 数据集上的检测性能。在零样本全模态大语言模型中，性能差异显著。MiniCPM-o-2.6 和 Phi-4-Multimodal 的表现中等，而 Qwen2.5-Omni-7B（Base）的准确率达到 59.83%。规模更大的 Qwen3-Omni-30B 教师模型进一步将性能提升至 66.95% 的准确率和 65.89% 的 F1 分数，展现了模型规模对零样本多模态推理的益处。对 Qwen2.5-Omni-7B 进行监督微调后，性能显著提升，准确率达到 68.72%，F1 分数达到 68.23%。这表明，经过筛选的推理轨迹能够提供有效的多模态监督，使模型与讽刺理解目标相契合。我们的最终框架 SarcasmMiner 取得了最佳表现，准确率达到 70.22%，F1 分数达到 70.23%。与监督微调相比，采用所提出奖励设计的强化学习在两项指标上均带来了稳定提升。值得注意的是，7B 版 SarcasmMiner 超越了所有零样本全模态模型，包括其 30B 教师模型，彰显了任务特定后训练的有效性。

4.1. 消融研究

表2呈现了与我们的三阶段框架相对应的全面消融研究。我们首先比较两种教师生成模式。与标准指令提示相比，启用思考模式后，准确率从 62.02% 提升至 66.95%，GAR 从 64.01 提升至 71.88。这些提升表明，思考模式能够生成质量更高、结构更清晰的推理轨迹，为下游蒸馏提供更有力的监督。接下来，我们考察构建 SFT 数据集时采用的不同初始化策略。从贪心解码切换为 Best-of-N 可同时提升性能和推理质量。采用多样化采样并进行筛选，可进一步将准确率提升至 68.72%，将 GAR 提升至 86.04，这表明在

表1： 在 *MUStARD++* 数据集上的检测结果。*Sarcas- mMiner*（*7B*）的表现始终优于其零样本骨干模型，也超过了包括 *30B*教师模型在内的更大型零样本基础模型。指标以百分比（%）表示。

模型	Acc	宏平均 F1
全模态大语言模型（零样本）		
MiniCPM-o-2.6	57.49	57.58
Phi-4-Multimodal	59.33	59.56
Qwen2.5-Omni-7B（Base）	59.83	59.83
Qwen3-Omni-30B（教师）	66.95	65.89
后训练（基于 *Qwen2.5-Omni-7B*）		
监督微调（多样化采样）68.72		68.23
SarcasmMiner（本文） **70.23**		**70.22**

表2：对我们三阶段框架的全面消融研究。我们评估了教师生成模式（阶段*1*）、用于监督微调初始化的轨迹选择策略（阶段*2*），以及在强化学习后训练中整合解耦奖励（阶段*3*）的影响。

配置	Acc	宏平均 F1	GAR
阶段 *1*：教师模式			
Qwen3-Omni-30B（指令版）	62.02	61.99	64.01
Qwen3-Omni-30B（思考版）	66.95	65.89	71.88
阶段 *2*：监督微调初始化			
监督微调（贪心）	65.42	65.14	77.57
监督微调（*N*选最优）	67.75	67.33	81.94
监督微调（多样化采样）	68.72	68.23	86.04
阶段 *3*：强化学习后训练			
GRPO（无监督微调）	60.29	60.11	66.17
监督微调+GRPO	69.96	69.82	84.47
监督微调+GRPO+GenRM（本文方法）	**70.23**	**70.22**	**90.43**

蒸馏过程中保持轨迹多样性，能够获得更强的学生模型初始化。最后，我们分析不同的强化学习后训练配置。直接应用标准强化学习而不进行 SFT（GRPO w/o SFT）会导致性能和推理质量下降（准确率为 60.29%，GAR 为 66.17），这反映出优化过程不稳定。加入 SFT 初始化后，训练稳定性显著提升（准确率为 69.96%，GAR 为 84.47）。进一步纳入 GenRM 奖励后，准确率（70.23%）和推理质量（GAR 为 90.43）均得到提升，表明显式考虑推理的奖励建模能够在策略优化过程中增强多模态落地能力。

4.2. 错误分析

为更好地理解解耦奖励机制带来的行为变化，我们比较了图2中三个代表性模型的混淆矩阵。

监督微调中的讽刺过度预测。 仅经过监督微调的后训练模型表现出强烈的讽刺预测偏向，导致大量假阳性（FP =0.45）。人工检查发现，该模型常出现“臆想式过度解读”现象，将中性话语过度解读为讽刺，凭空捏造不存在的语用冲突，

不受视听语境的支持。

In标准 GRPO 下的稳定性。应用 GR PO

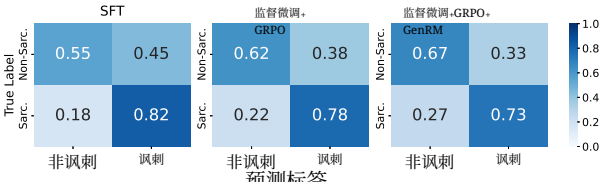

图2： *MUStARD++* 数据集上不同后训练方法的按行归一化混淆矩阵。

没有 GenRM 奖励时，模型会优化准确率，但缺乏对推理质量的约束。虽然假阳性从 0.45 略微降至 0.38，但假阴性（FNs）却从 0.18 升至 0.22。策略以牺牲对真正讽刺的敏感度为代价，换取非讽刺检测上的微小提升，反映出无约束探索的不稳定性。

使用 SarcasmMiner 改进锚定。 在 GenRM 的逻辑惩罚引导下，SarcasmMiner 展现出更加保守且更为均衡的预测行为。假阳性降至 0.33，表明多模态幻觉有所减少，而真阴性增至 0.67。此外，尽管真阳性的绝对数量略有下降，SarcasmMiner 仍取得了更高的召回率（0.70）。这些结果印证了 GAR 的改进，表明 SarcasmMiner 能够有效学习仅在有具体且逻辑合理的视听证据支持时才做出讽刺预测。

案例研究。 为直观展示我们的框架如何缓解幻觉，我们分析了一个非讽刺样本：“所以，今年春天，我可以去国际空间站。哦，天哪，这是一次通往天堂的旅行。”SFT 模型过度依赖文本夸张，而忽略了视听语境：“‘哦，天哪……’这句话以夸张、近乎戏剧化的热情说出。‘一次通往天堂的旅行’对于一件平凡的事件而言是夸张说法，这是讽刺的典型特征。”这些解读表明，模型虚构了一种嘲弄性的声学语调，并荒谬地将太空旅行归类为平凡事件。相比之下，SarcasmMiner 将其推理建立在细粒度的多模态证据之上：“说话者的语调高亢、充满活力，并洋溢着真诚的热情。从视觉上看，*Penny* 睁大双眼、张开嘴巴并露出灿烂的笑容——这些都是真诚惊叹的明确表现。 通过明确将文本夸张与真实的声学和视觉情绪状态相结合，SarcasmMiner 展现出了有效的跨模态锚定能力。

5. 讨论

我们的实验表明，后训练在使多模态基础模型适应讽刺推理方面发挥着核心作用。尽管零样本全模态模型表现尚可，但它们往往无法始终将预测可靠地锚定于语音韵律和面部表情。通过明确向模型展示结构化的跨模态解释，监督式推理蒸馏能够提升这种对齐效果。此外，在强化学习期间仅优化预测准确率，可能会诱发捷径行为，即模型在缺乏可靠多模态依据的情况下给出正确标签。针对推理有效性设置解耦奖励，可鼓励模型连贯地整合文本、声学和视觉信号，而非仅仅优化标签正确性。这些发现表明，有效的基础模型后训练应结合结构化推理监督与明确的奖励约束，尤其适用于涉及细微跨模态不一致的任务。

6. 参考文献

[1] A. Radford、J. W. Kim、T. Xu、G. Brockman、C. McLeavey 和 I. Sutskever，“通过大规模弱监督实现鲁棒语音识别”，载于国际机器学习会议。PMLR，2023年，第28 492–28 518页。[2] A. Hurst、A. Lerer、A. P. Goucher、A. Perelman、A. Ramesh、A. Clark、A. Ostrow、A. Welihinda、A. Hayes、A. Radford等，“GPT-4o系统卡”，*arXiv*预印本 *arXiv:2410.21276*，2024年。[3] Y. Chu、J. Xu、X. Zhou、Q. Yang、S. Zhang、Z. Yan、C. Zhou 和 J. Zhou，“Qwen-Audio：通过统一的大规模音频-语言模型推进通用音频理解”，*arXiv*预印本 *arXiv:2311.07919*，2023年。[4] J. Schulman、F. Wolski、P. Dhariwal、A. Radford 和 O. Klimov，“近端策略优化算法”，*arXiv*预印本 *arXiv:1707.06347*，2017年。[5] L. Ouyang、J. Wu、X. Jiang、D. Almeida、C. Wainwright、P. Mishkin、C. Zhang、S. Agarwal、K. Slama、A. Ray等，“通过人类反馈训练语言模型遵循指令”，神经信息处理系统进展，第35卷，第27 730–27 744页，2022年。[6] Z. Shao、P. Wang、Q. Zhu、R. Xu、J. Song、X. Bi、H. Zhang、M. Zhang、Y. Li、Y. Wu等，“DeepSeekMath：在开放语言模型中拓展数学推理的极限”，*arXiv*预印本 *arXiv:2402.03300*，2024年。[7] R. Schifanella、P. De Juan、J. Tetreault 和 L. Cao，“检测多模态社交平台中的讽刺”，载于第*24*届*ACM*国际多媒体会议论文集，2016年，第1136–1145页。[8] J. Wei、X. Wang、D. Schuurmans、M. Bosma、F. Xia、E. Chi、Q. V. Le、D. Zhou等，“思维链提示能够激发大语言模型的推理能力”，神经信息处理系统进展，第35卷，第24 824–24 837页，2022年。[9] T. Kojima、S. S. Gu、M. Reid、Y. Matsuo 和 Y. Iwasawa，“大语言模型是零样本推理器”，神经信息处理系统进展，第35卷，第22 199–22 213页，2022年。[10] L. Huang、W. Yu、W. Ma、W. Zhong、Z. Feng、H. Wang、Q. Chen、W. Peng、X. Feng、B. Qin等，“大语言模型中的幻觉综述：原理、分类体系、挑战与开放问题”，*ACM*信息系统汇刊，第43卷，第2期，第1–55页，2025年。[11] S. Castro、D. Hazarika、V. Pˊerez-Rosas、R. Zimmermann、R. Mi-halcea 和 S. Poria，“迈向多模态讽刺检测（显然是一篇完美的论文）”，载于第*57*届计算语言学协会年会论文集，A. Ko-r̄honen、D. Traum 和 L. M`arquez 编。意大利佛罗伦萨：计算语言学协会，2019年7月，第4619–4629页。[12] A. Ray、S. Mishra、A. Nunna 和 P. Bhattacharyya，“用于讽刺情感识别的多模态语料库”，载于第十三届语言资源与评测会议论文集，法国马赛，2022年6月，第6992–7003页。[13] D. Raghuvanshi、X. Gao、Z. Li、S. Bansal、M. Coler、N. Kumar 和 S. Nayak，“利用图注意力网络进行多模态讽刺检测的模态内关系与情感不一致学习”，载于*ICASSP 2025——2025*年*IEEE*声学、语音与信号处理国际会议（*ICASSP*）。IEEE，2025年，第1–5页。[14] X. Gao、S. Bansal、K. Gowda、Z. Li、S. Nayak、N. Kumar 和 M. Coler，“AMUSED：一种结合双模态数据增强、用于多模态讽刺检测的注意力深度神经网络”，*arXiv*预印本 *arXiv:2412.10103*，2024年。[15] Q. Yang、L. Yang、J. Zeng、K. Ma 和 H. Lin，“Sarcasm-R1：通过聚焦推理增强讽刺检测”，载于计算语言学协会论文集：*EMNLP 2025*精选论文，2025年，第10 773–10 785页。[16] A. Saha、V. Suresh、T. Hospedales 和 V. Demberg，“MUStReason：用于诊断视频大语言模型在多模态讽刺检测中语用推理能力的基准”，*arXiv*预印本 *arXiv:2510 23727*，2025年.。

[17] D. Wang、S. Liu、T. Zhang、Y. Chen、J. Li 和 H. Meng，“Emotionthinker：面向可解释语音情感推理的韵律感知强化学习”，*ICLR*，2026。[18] Y. Chen、S. Yang、Y. Xiang、R. Song、Y. Huang 和 Z. Yu，“GRPO引导的模态选择增强的 LoRA 微调大语言模型用于多模态情感识别”，见计算语言学协会论文集：*EMNLP 2025*，2025，第 19458–19471 页。[19] Y. Gao、S. Zhang、R. Yang、Z. Cui 和 J. Feng，“B-GRPO：基于批组相对策略优化的无监督语音情感识别”，*2026* 年 *IEEE* 声学、语音与信号处理国际会议（*ICASSP*），2026。[20] S. Arora、K.-W. Chang、C.-M. Chien、Y. Peng、H. Wu、Y. Adi、E. Dupoux、H.-Y. Lee、K. Livescu 和 S. Watanabe，“口语语言模型图景：一项全面综述”，*arXiv* 预印本 *arXiv:2504.08528*，2025。[21] Y. Yao、T. Yu、A. Zhang、C. Wang、J. Cui、H. Zhu、T. Cai、H. Li、W. Zhao、Z. He等，“MiniCPM-V：手机上的 GPT-4V 级多模态大语言模型”，*arXiv* 预印本 *arXiv:2408.01800*，2024。[22] J. Xu、Z. Guo、J. He、H. Hu、T. He、S. Bai、K. Chen、J. Wang、Y. Fan、K. Dang、B. Zhang、X. Wang、Y. Chu 和 J. Lin，“Qwen2.5-Omni 技术报告”，*arXiv* 预印本 *arXiv:2503.20215*，2025。[23] J. Xu、Z. Guo、H. Hu、Y. Chu、X. Wang、J. He、Y. Wang、X. Shi、T. He、X. Zhu等，“Qwen3-Omni 技术报告”，*arXiv* 预印本 *arXiv:2509.17765*，2025。[24] Y. Zhao、J. Huang、J. Hu、X. Wang、Y. Mao、D. Zhang、Z. Jiang、Z. Wu、B. Ai、A. Wang、W. Zhou 和 Y. Chen，“SWIFT：用于微调的可扩展轻量级基础设施”，2024。[在线]。可获取：https://arxiv.org/abs/2408.05517[25] J. Rasley、S. Rajbhandari、O. Ruwase 和 Y. He，“DeepSpeed：系统优化使训练超过 1000 亿参数的深度学习模型成为可能”，见第*26*届*ACM SIGKDD* 知识发现与数据挖掘国际会议论文集，2020，第3505–3506 页。[26] E. J. Hu、Y. Shen、P. Wallis、Z. Allen-Zhu、Y. Li、W. Wang 和 Z. Chen，“LoRA：大语言模型的低秩适配”，见学习表征国际会议（*ICLR*），2022。