思维模型能否通过思考检测仇恨模因?

Mohamed Bayan Kmainasi 卡塔尔大学 卡塔尔多哈
mk2314890@qu.edu.qa

Mucahid Kutlu Ali Ezzat Shahroor 卡塔尔大学 卡塔尔计算研究所 卡塔尔多哈 卡塔尔多哈 mucahidkutlu@qu.edu.qa
alsh34060@hbku.edu.qa

Abul Hasnat
APAVI.AI 法国
Blackbird.AI 美国
m
hasnat@gmail.com

Firoj Alam 卡塔尔计算研究所
卡塔尔多哈 fialam@hbku.edu.qa

arXiv:2603.01225v1 [cs.CL] 1 Mar 2026

摘要

仇恨模因通常需要组合式多模态推理：图像和文本单独来看可能无害，但两者的交互会传达有害意图。尽管基于思考的多模态大语言模型（MLLMs）近年来推动了视觉-语言理解的发展，但其在仇恨梗图分析方面的能力仍未得到充分探索。我们提出一种基于强化学习的后训练框架，通过任务特定奖励和一种新颖的组相对策略优化（GRPO）目标，提升基于思考的多模态大语言模型的推理能力。具体而言，我们（*i*）对现成的多模态大语言模型在仇恨梗图理解方面进行系统的实证研究，（*ii*）通过蒸馏生成弱监督/伪监督的思维链（CoT）推理依据，以扩展现有的仇恨梗图数据集，并（*iii*）提出一种基于 GRPO 的目标，同时优化梗图分类和解释质量，以促进细粒度逐步推理。在仇恨模因基准上的实验表明，我们的方法达到了最先进的结果，准确率和 F1 提升约 1%，解释质量提升约 3%。我们将公开发布代码、数据扩展和评估资源，以支持结果复现。[1]

CCS 概念

• **计算方法** → **计算机视觉问题**。

关键词

仇恨模因检测；多模态大语言模型；强化学习；多模态推理

ACM 引用格式：
Mohamed Bayan Kmainasi、Mucahid Kutlu、Ali Ezzat Shahroor、Abul Hasnat 和 Firoj Alam。2026年。《思维模型能否通过思考检测仇恨模因? 》。载于*ACM Web Conference 2026* 会议伴随论文集（*WWW Companion '26*），*2026*年*4*月*13*日至*17*日，阿拉伯联合酋长国迪拜。ACM，纽约，美国，10页。https://doi.org/10.1145/3774905.3795465

[1]https://github.com/MohamedBayan/MemeReason

本作品依据知识共享署名 4.0 国际许可协议授权。*WWW Companion '26*，迪拜，阿拉伯联合酋长国© 2026年版权所有者/作者所有。ACM ISBN 979-8-4007-2308-7/2026/04 https://doi.org/10.1145/3774905.3795465

1 引言

模因将图像和文本与幽默及文化典故相结合，是典型的多模态内容。它们已成为社交媒体上的主要传播形式，迅速塑造着集体话语 [1, 40]。虽然模因往往无害，但也具有双重性，可能被利用来传播仇恨言论、虚假信息和宣传 [47]。仇恨模因运用幽默和反讽，可能会淡化有害内容，进而使敌对观点逐渐常态化 [40, 44]。此外，模因还是强大的政治传播工具，能够通过简单且富有情感共鸣的视觉内容有效传达复杂思想，从而实现说服和操纵 [1, 37]。鉴于社交媒体无处不在，及时准确地检测和审核仇恨内容，对于营造更安全的网络环境至关重要。受这些挑战的启发，本研究旨在借助多模态大语言模型（*MLLMs*）的最新进展，改进仇恨模因检测。

近年来，多模态大语言模型在视觉-语言理解方面取得了显著进展，并在广泛的多模态任务中展现出强大的泛化能力 [49, 53]。除具备出色的表征学习能力外，这些模型还日益展现出指令遵循、跨模态定位以及向仅有有限特定任务监督的任务进行稳健迁移等新兴能力，因此成为应对复杂内容理解问题的理想基础模型。基于思维的多模态大语言模型 代表了多模态推理的最新进展，通过有意进行多步推理，展现出更强的整合视觉与文本信息的能力 [56]。这一范式与模因尤其相关，因为模因的含义通常是隐含的，并且依赖图像与文本之间的交互，而非单模态线索。这类模型的训练强调思维链（CoT）监督，即对中间推理步骤进行显式建模，并进一步通过强化学习（RL）技术进行优化 [14]。

通过多种方法，模因理解已取得显著进展，包括智能体推理框架 [30, 31], 、少样本适应 [11, 18], 以及在多模态数据集 [2, 24]上进行监督微调（SFT）。然而，直到最近才有研究开始探索将思维链、监督训练和强化学习用于模因分析。此外，基于思维的多模态大语言模型尚未用于仇恨梗图分析。

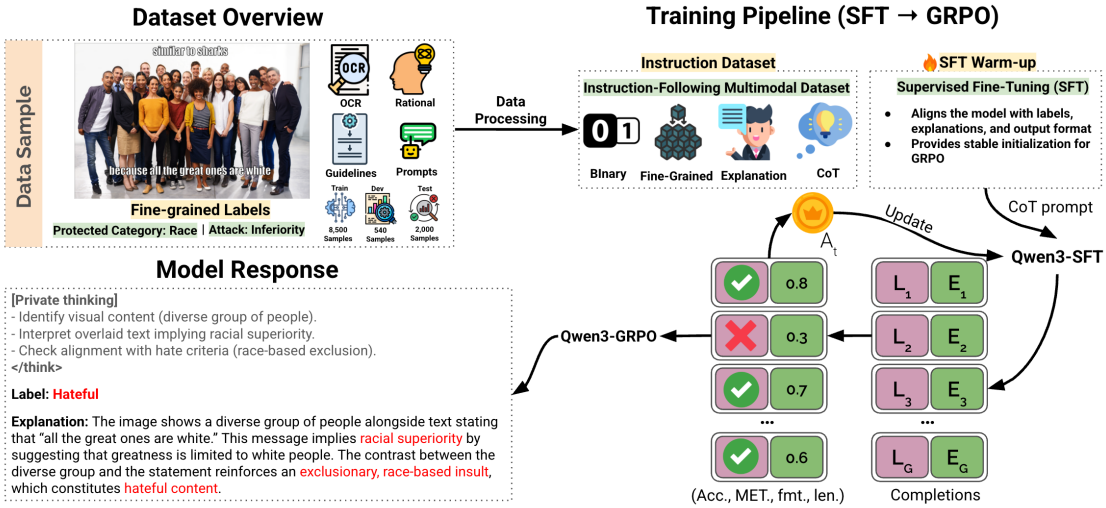


图 1：所提出的仇恨梗图分析框架概览。从原始梗图出发，我们提取二元和细粒度监督信息（受保护类别和攻击类型）、OCR文本及指南，以构建指令遵循数据集。我们利用强大的多模态大语言模型提供的弱监督，蒸馏出逐步展开的思维链推理依据。模型通过两阶段后训练流程进行训练：先进行监督微调预热，再进行基于GRPO的强化学习，共同优化分类准确率和解释质量。在推理阶段，系统生成多个候选标签—解释对并对其评分，以选出最终预测。缩写：Acc. = 准确率；MET = METEOR；Len = 解释长度；Fmt =格式合规性。

为弥补这一研究空白，如图 1 所示，我们研究了这一方向，并提出了一种基于思维链监督推理范式、通过强化学习训练的方法。具体而言，我们首先使用预训练思考型多模态大语言模型研究并分析思维链对模因分析的影响，随后提出一种新方法（详情见第 *4* 节），对基于思维的多模态大语言模型进行微调，以用于仇恨梗图分析。

我们的贡献总结如下：

- **思维链推理的实证研究：**我们在零样本和思维链提示设置下，评估近期预训练的多模态大语言模型对仇恨梗图的理解能力，并联合考察分类和解释生成，以确定一个性能强劲的基础模型。
- **扩展逐步推理：**我们扩展了现有的仇恨梗图数据集，加入蒸馏的逐步思维链推理依据；这些依据通过强大的商业多模态大语言模型进行弱监督／伪监督获得。
- **GRPO优化目标：**我们为梗图分类引入一种基于GRPO的目标，并采用基于METEOR的奖励目标，联合优化仇恨梗图分析中的标签正确性和解释质量。
- **两阶段GRPO微调：**我们设计了一套后训练流程，先进行监督微调预热，再采用基于GRPO的强化学习，实现可解释的仇恨梗图检测。

我们的发现可总结如下：*(a)* 基于思考的预训练多模态大语言模型始终能从思维链提示中获益；*(b)*在基于GRPO的后训练中，加入细粒度监督和蒸馏后的思维链推理依据，可进一步提升分类和解释质量；以及*(c)*所提出的GRPO目标函数和基于强化学习的训练方法，在仇恨模因分类和解释方面达到了最先进的结果。

2 相关工作

2.1 多模态仇恨模因检测

早期的模因检测系统依赖单模态线索，例如基于OCR的文本特征与传统机器学习算法相结合 [3, 7]，基于CNN的图像分类器和图文融合方法 [48]。然而，这些方法往往难以捕捉复杂图文交互所产生的微妙且依赖语境的含义。

现有关于模因审核的研究大多聚焦于二元分类，区分仇恨内容和非仇恨内容等类别 [9]。为支持这类研究，研究者构建了多个基准数据集，包括 Kiela 等人推出的仇恨模因数据集 [23]，以及针对新冠肺炎和美国政治模因的 HarMeme [41]。

只有少数研究探索了细粒度标注。例如，Mathias 等人 [34]根据受保护类别和攻击类型，为仇恨模因重新标注了细粒度属性；Fersini 等人 [16] 则推出了 MAMI，这是一个用于厌女模因检测的数据集，可根据表达的厌女类型进行二元分类和子类型级别的分类。

近年来，多项研究已从检测进一步拓展到可解释性、分类标注以及增强的多模态推理。例如，Grasso 等人 [17] 利用 ConceptNet 进行知识图谱推理，以增强多模态分类并丰富语境理解。Mod-Hate 框架 [11]通过模块化大语言模型组件来应对少样本仇恨模因检测，而 Lin 等人 [28] 提出了一种多模态辩论框架，其中多个大语言模型智能体在做出最终决策之前交换推理依据。在一项相关工作中，Kmainasi 等人

[24] 通过纳入人工验证的理由扩展了 FHM，从而实现对可解释多模态大语言模型的监督训练。

尽管取得了这些进展，关于如何增强多模态输入上的思维链推理，系统性研究仍然有限。现有研究已在特定场景中探索思维链提示或结构化推理，例如厌女模因分类中的多跳推理 [25], 多模态隐喻检测 [55], 以及用于有害梗图检测的蒸馏式溯因推理 [29]。然而，利用强化学习/GRPO进行多模态仇恨模因检测仍鲜有研究。据我们所知，只有近期或同期的少数研究 [36] 开始探讨在模因审核中应用GRPO。以往研究主要依赖监督式或基于提示的方法，尚未充分考察强化学习在提升该领域多模态推理能力方面的潜力。

2.2 多模态大语言模型的强化学习

强化学习是机器学习的核心范式，智能体通过与环境交互并获得奖励来学习做出决策 [20]。多模态大语言模型的最新进展已将强化学习用于动态反馈下的后训练，奖励信号由此引导模型行为 [54]。基于人类反馈的强化学习（RLHF）等方法已被广泛用于对齐和优化大语言模型的行为 [21, 39]。

这些方法通常采用近端策略优化（PPO） [45] 和直接偏好优化（DPO） [42] 等优化算法，调整模型参数以最大化期望奖励信号。最近，DeepSeek-R1 模型引入了 GRPO [14],，实现了最先进水平的推理性能。与依赖已学习评估模型的传统 RLHF 方法不同，GRPO 采用基于规则的奖励，明确鼓励生成高质量的思维链，并倾向于选择推理质量更高的轨迹。GRPO 已有效应用于多个领域，包括数学推理 [46], 、自训练 [43], 和代码生成 [12]。在本研究中，我们通过优化基于思维的多模态大语言模型，将 GRPO 应用于仇恨梗图分析，并研究思维链监督与基于强化学习的后训练如何相互作用，从而提升检测和解释效果。

3 数据集

3.1 数据集准备

对于仇恨模因分析，大多数先前研究主要聚焦于粗粒度分类 [9]。近期研究表明，使用更详细的监督信息训练模型可以显著提升任务表现，因为这类信息能够模拟人类在分析多模态输入并作出最终判断时的逐步推理过程 [60]。有效训练基于思维的多模态大语言模型，需要提供超越二元标签的充足且结构化的任务专属信息。因此，我们整理了包含二元标签和细粒度标签的现有数据集，并扩展了原始仇恨模因数据集，为分配的标签添加解释和推理依据，以及逐步思维链推理轨迹。

3.1.1 仇恨模因——原始数据集。仇恨模因 [22] 是多模态仇恨言论检测领域应用最广泛的早期基准数据集之一，包含约 12k 个结合文本与视觉内容的模因。该数据集经过精心整理，确保有效分类需要对两种模态进行联合多模态理解，因为仅凭文本或图像都不足以可靠地判定标签。数据集包含合成生成的表情包和真实案例，并在仇恨内容与非仇恨内容之间保持均衡分布。[2] 表1总结了数据集分布。

类别标签	训练集	开发集	Test	总计
非仇恨	5,481	340	1,250	7,071
仇恨	3,019	200	750	3,969
总计	**8,500**	**540**	**2,000**	**11,040**

表1：数据集中标签的分布。

3.1.2 解释/推理。 我们利用近期发布的 MemeIntel 工作 [24], 中引入的推理依据，这些推理依据提供了与二元标签相对应的经人工核验的解释。

3.1.3 细粒度标签。 为获取原始数据集的细粒度标注，我们利用 Mathias 等人 [33], 公开提供的标注，这些标注根据受保护类别（宗教、种族、性别、残障和国籍）以及攻击类型（非人化、低劣化、煽动暴力、嘲弄、蔑视、侮辱性称呼和排斥）对每个仇恨实例进行分类。

3.2 逐步推理——我们的扩展

该数据集的最终扩展引入了逐步推理依据，这些依据对于基于思考的多模态大语言模型训练至关重要。然而，从人工标注者处收集高质量推理轨迹是一项极为耗时且劳动密集的过程。与此同时，商用推理型模型的最新进展已在实践中展现出强大而可靠的推理能力 [13, 26]。受这些观察的启发，我们采用基于蒸馏的方法来生成逐步推理轨迹，该方法称为思维链蒸馏（CoTD）。

我们使用 GPT-4.1，根据模因图像、提取文本、标注指南以及二元标签和细粒度标签，生成逐步推理轨迹。这些私有推理序列仅用于在训练期间监督多模态大语言模型，不会在推理期间公开。我们的方法遵循了此前利用 CoTD 提升模型推理性能的研究 [60]。为防止标签泄漏，我们明确提示模型独立推理，不得复制或改写参考解释。提示模板已发布在 GitHub 代码仓库中。

使用大语言模型评判评估思维链。 我们参照 Kmainasi 等人的研究 [24]，采用大语言模型评判协议评估 CoTD 的质量。我们使用强大的视觉语言模型作为自动评判器，以实现可扩展且一致的评估。每个样本均从四个维度进行评估：信息丰富度，衡量对显著视觉和文本证据的运用；清晰度，评估逻辑结构和易于理解的程度；合理性，

[2] 在这项工作中，我们关注未见过的数据划分，因为这些数据在 [24] 中补充了说明。

指标	InternVL3.5	Phi-3.5	一致性
忠实度	**4.782**	4.557	0.944
清晰度	**4.733**	4.537	0.943
合理性	**4.608**	4.241	0.934
信息丰富度	**4.399**	4.097	0.942
平均值	**4.631**	4.358	0.940

表 2：训练集+开发集的平均大语言模型评判得分（1–5）和一致性指标。InternVL3.5 =OpenGVLab InternVL3.5- 8B，Phi-3.5 = Phi-3.5-Vision-Instruct。一致性表示评审者间一致性。

评估推理是否提供了连贯且符合人类判断的解释；以及忠实度，衡量其是否以可观察到的表情包内容为依据，且不包含幻觉或无依据的主张。

我们使用两个强大的视觉语言模型，**InternVL3.5** 和**Phi-3.5**，作为独立评审者，并使用 5 点李克特量表对每个维度进行评分。表 2 报告了所得分数以及评审者间一致性。序数评分的一致性使用$r^*_{wg(j)}$ 指数 [19], 进行量化，该指数通过比较观测评分方差与完全不一致时的最大方差来衡量一致性:

$$r^*_{wg(j)} = 1 - \frac{S_X^2}{\sigma^2_{\text{mv}}},$$

其中， S_X^2 表示评审者之间的方差， σ^2_{mv} 表示 5 点量表下可能的最大方差。数值越高，表示共识越强。如表 2 所示，所有维度的一致性得分均稳定地保持在较高水平，表明基于大语言模型的判断可靠且稳定。完整的评估材料已在补充材料中发布。

两位评审者给出的平均分均高于 4.0，表明整体解释质量很高。忠实度 和 清晰度 在 InternVL3.5 和 Phi-3.5 的评分中均为最高，这表明生成的推理轨迹以可观察到的表情包内容为依据，逻辑结构清晰且易于理解。

4 方法论

本节介绍我们提出的基于强化学习的方法，该方法使用基于思维的多模态大语言模型进行仇恨模因分析，如图 1 所示。我们的方法旨在通过结构化输出格式和多阶段训练流程，联合训练多模态大语言模型，使其能够执行分类、解释生成和逐步推理。

框架概述。 如图 1 所示，我们提出了一种端到端仇恨梗图分析框架，整合了细粒度监督、思维链推理和基于强化学习的后训练。对于由图像和嵌入文本组成的原始表情包，我们首先结合多种监督信号构建指令跟随训练示例，包括仇恨／非仇恨二元标签、标明受保护类别和攻击类型的细粒度标注、OCR提取文本以及分类指南。这种统一表示促使模型综合视觉和文本线索进行推理，并以既有的仇恨分类体系为依据生成预测。

为实现逐步推理，我们采用弱监督思维链蒸馏策略，使用强大的商业多模态大语言模型，根据表情包内容、标签和指南生成私有推理轨迹。这些蒸馏后的思维链推理依据仅用于训练。模型优化采用两阶段后训练流程。首先，通过监督微调使模型适应结构化输出格式、金标准标签、解释和蒸馏思维链监督。其次，GRPO通过针对每个表情包采样多个候选项（私有思维链 + 标签 + 解释／推理依据），进一步优化模型，并强化相对奖励更高的候选项，从而共同优化标签正确性、格式合规性和解释质量。在推理阶段，模型进行内部推理，并输出预测标签及其支持性解释。

在接下来的小节中，我们首先对任务进行形式化定义。然后，我们分多个步骤介绍多模态大语言模型的优化流程。最后，我们详细说明训练与评估的设置。

4.1 任务定义

给定一个由图像和提取文本组成的模因 x，任务是预测该模因是否具有仇恨性，并生成自然语言推理依据来证明这一判断。形式上，给定一个指令式提示 $c(x)$，模型会生成如下形式的结构化输出 y：

$$y = \big\langle \texttt{<think>}\ t\ \texttt{</think>}\ \texttt{Label:}\ \hat{\ell}\ \texttt{Explanation:}\ \hat{e} \big\rangle,$$

其中， t 表示私有思维链推理轨迹，$\hat{\ell}$ 表示预测类别标签，$\hat{e}$ 是用于证明预测标签的自然语言解释。令 D 表示带标签数据集，如下所示:

$$\mathcal{D} = \{(x_i, y_i^\star)\}_{i=1}^{N},$$

其中 $y_i^\star$包含标准标签、解释和蒸馏后的教师推理。 i 表示样本编号，N 表示 D中的样本总数。

4.2 基于思维的多模态大语言模型训练

我们的训练流程包含两个步骤：首先通过监督微调初始化多模态大语言模型的参数，然后应用 GRPO。下面简要介绍训练流程的细节。

4.2.1 监督微调预热。我们通过在 D 上进行监督微调来初始化多模态大语言模型。令π_θ 表示由θ参数化的条件自回归模型。给定一个指令式提示 c（x），其中编码了模因图像和 OCR提取文本，模型经过训练后生成结构化目标输出 $y^\star$。通过最小化负对数似然损失来优化参数 θ：

$$\mathcal{L}_{\text{SFT}}(\theta) = \mathbb{E}_{(c,y^\star)\sim\mathcal{D}}\left[-\log \pi_\theta(y^\star \mid c)\right] \tag{1}$$

$\mathbb{E}_{(c,y^\star)\sim\mathcal{D}}$ 表示训练集上的经验期望（通过小批量平均值近似），即我们对以提示c为条件的目标输出 $y^\star$ 的负对数似然取平均。此阶段使模型与标准标签、解释和蒸馏后的推理轨迹保持一致，为后续基于 GRPO 的优化提供强有力的初始化。

多模态大语言模型初始化后，我们根据留出验证集上的评估损失选择最佳检查点（用于下一步）。

*4.2.2 GRPO*优化。 GRPO通过对比每个输入对应的多个采样输出，并强化优于同组其他输出的推理轨迹来优化模型，从而实现稳定有效的以推理为中心的后训练。它能优化多模态大语言模型，提升其推理能力和解释质量。

在此阶段，多模态大语言模型被视为随机策略 π_θ，根据提示 $c(x)$ 生成输出。对于每个输入，我们从 π_θ 中抽取一组包含 K 个候选输出的样本 $\{y_1, \ldots\ldots, y_K\}$，并为每个补全计算一个标量奖励。总体奖励函数定义为多个组成部分的加权组合：

$$R(y) = \alpha_{\text{fmt}} R_{\text{fmt}}(y) + \alpha_{\text{lbl}} R_{\text{lbl}}(y) + \alpha_{\text{len}} R_{\text{len}}(y) + \alpha_{\text{met}} R_{\text{met}}(y), \tag{2}$$

其权重为 $\alpha_{\text{fmt}} = 0.5$、$\alpha_{\text{lbl}} = 0.4$、$\alpha_{\text{len}} = 0.05$ 和 $\alpha_{\text{met}} = 0.05$，通过经验设定。$R(y)$ 的各个组成部分如下：

- Rfmt(y)（**格式**）：鼓励结构一致的输出，并明确包含推理、预测和解释部分。
- Rlbl(y)（**标签正确性**）：根据金标准标注奖励正确类别的选择，确保推理以正确决策为依据。
- Rlen(y)（**长度正则化**）：以软约束的方式鼓励推理依据接近目标长度（约 100 个词，与 [24] gold rationales 匹配），避免解释过短或过于冗长：$R\text{len}(y) = \exp\left(-\frac{(L-100)^2}{2\sigma^2}\right)$。我们选择 $\sigma = 20$，使长度奖励充当软正则项，温和地抑制偏离目标长度的极端情况，而不会主导 GRPO目标函数。
- Rmet(y)（**语义相似度**）：通过 METEOR分数奖励与 gold rationale 在语义上接近的解释 [5],[3]，在标签准确率之外鼓励生成质量更高的论证。

目标函数。 GRPO 使用组平均奖励 $\bar{R}$ 作为基线来计算归一化优势：

$$\bar{R} = \frac{1}{K}\sum_{k=1}^{K} R_k, \qquad A_k = \frac{R_k - \bar{R}}{\text{std}(\{R_j\}) + \varepsilon}$$

设 $\pi_{\theta_{\text{old}}}$ 表示更新前的策略。对于每个生成的词元 $a_{k,\,t}$ 在状态 $h_{k,\,t}$下，重要性比例如下。

$$r_{k,t} = \frac{\pi_\theta(a_{k,t} \mid h_{k,t})}{\pi_{\theta_{\text{old}}}(a_{k,t} \mid h_{k,t})}$$

GRPO目标函数通过KL 正则化最大化裁剪代理目标：

$$\mathcal{J}_{\text{G}}(\theta) = \mathbb{E}\left[\sum_{k=1}^{K}\sum_{t} \min\Big(r_{k,t}A_k, \text{clip}(r_{k,t}, 1-\epsilon, 1+\epsilon)A_k\Big) - \beta\,\text{KL}\big(\pi_\theta(\cdot \mid h_{k,t}) \,\|\, \pi_{\text{ref}}(\cdot \mid h_{k,t})\big)\right] \tag{3}$$

其中，ϵ 表示 PPO 风格的裁剪阈值，β 控制 KL 正则化项的贡献，而 π_{ref} 是从监督微调检查点初始化的固定参考策略。KL 项约束更新后的策略，使其与参考模型保持接近，从而提高训练稳定性并防止奖励过度优化。

总体而言，该优化鼓励更高奖励的推理轨迹，同时通过裁剪和 KL 正则化稳定更新。

5 实验设置

5.1 模型

在我们的实验中，我们同时使用开放权重和封闭权重的多模态大语言模型。我们的开放权重基线包括Llama-3.2-11B、Llama-4-Scout-17B [15], Qwen3-VL-8B [27]以及Gemma-3-12B [50]。作为代表性的封闭权重模型，我们使用GPT-4。1 [38]。在思维链实验中，我们使用Qwen3-VL-8B-Thinking [4], ，这是随Qwen3-VL系列发布的推理增强检查点，可在开放权重设置下进行受控的思维链消融，并在 [27] 中被报告于MMMU [58] 和MathVista [32]等多模态推理基准上表现强劲。总体而言，我们选择这些模型，是因为其技术报告和模型卡显示它们在既有视觉语言基准上取得了优异结果，使其成为多模态理解任务中具有竞争力的基线。我们的实验同时包括思考模型和指令微调模型。

5.2 零样本与思维链推理

我们在零样本和思维链提示设置下，对上述开源权重和封闭权重的多模态大语言模型进行基准测试，其中包括经过指令微调的模型和显式“思考”变体。为确保公平，在非思维链设置下，我们会关闭支持思考的模型的内部思考模式。此设置用于评估预训练多模态大语言模型在未经任务特定微调时的推理能力，并作为参考，帮助我们了解单靠通用多模态推理能够达到怎样的性能。

5.3 监督微调预热

先前研究表明，在基于强化学习的优化之前进行监督微调，能够提高面向推理的大语言模型的训练稳定性和下游性能 [46, 51]。受此发现启发，我们在 GRPO 之前引入监督微调预热阶段，以注入任务特定知识，并初步建立进行细粒度仇恨模因分析所需的结构化推理格式。

不含显式思维链的监督微调。 我们考虑两种不包含显式思维链监督的监督微调变体。第一种变体使用二元分类标签（仇恨与非仇恨）、细粒度标注（受保护类别和攻击类型）以及人工撰写的解释混合训练，记为**监督微调预热（分类 + 反馈生成 +解释，无思维链）**。第二种变体不使用细粒度标签，仅使用二元标签和解释进行训练，记为**监督微调预热（分类 + 解释，无思维链）**。在这两种设置中，模型均以预训练思考型骨干模型为初始化。为保留其推理能力并保持输出格式一致性，训练时会在助手输出前缀中加入空的 <think></think>标签，并在计算损失时将其排除。

[3]我们选择METEOR，因为它可以在训练过程中高效计算。

使用蒸馏思维链的监督微调。 为研究监督微调期间显式推理监督的影响，我们进一步使用 GPT-4.1 生成的逐步推理轨迹对**监督微调预热（分类 +反馈生成 + 解释，无思维链）**检查点进行微调，并将这些轨迹置于 <think>...</think> 块中（见第 3.2 节）。该变体记为**监督微调预热（分类 + 反馈生成 +解释，CoTD）**；它在保持相同输出结构的同时，使模型在监督微调期间接触细粒度推理信号。尽管某些监督微调变体不包含显式思维链监督，但得益于预训练思考型初始化，模型在推理时仍能生成内部推理。总体而言，监督微调预热阶段主要使模型与任务监督和输出结构保持一致，而 GRPO 则进一步提升推理能力和解释质量。

5.3.1 监督*GRPO*。 我们使用多个初始化检查点应用监督GRPO，以分析不同监督微调预热策略的效果。作为基线，我们直接从预训练的基于思考的骨干模型初始化GRPO，记为**Qwen3-VL-8B-思考 + GRPO**。此外，我们还从每种监督微调预热变体初始化GRPO：**监督微调预热（分类 + 反馈生成 + 解释，无思维链）**、**监督微调预热（分类 + 解释，无思维链）**，以及**监督微调预热（分类 + 反馈生成 + 解释，蒸馏思维链）**。所有模型均在带标签的仇恨表情包数据集上使用监督GRPO进行优化。如第4.2.2节所述，GRPO针对每个输入采样 K 个候选补全，并使用复合奖励函数更新策略；该函数促进结构有效的输出、正确预测和高质量解释，同时支持稳定优化。

5.4 训练设置

所有模型均使用 DeepSpeed ZeRO-3，在 4 张 NVIDIA H200 GPU 上进行全参数优化训练。SFT 训练 3 个轮次，根据开发集损失选择最佳检查点，并在每个轮次后进行评估。**SFT（分类 + 反馈生成+解释，蒸馏思维链）**[4] 模型从相应的**SFT（分类 + 反馈生成 + 解释，无思维链）**检查点初始化，并进一步微调 3 个轮次。对于 SFT，每设备批次大小设为 4，梯度累积步数固定为 1。使用 AdamW 进行优化，$\beta_1 = 0.9, \beta_2 = 0.95, \epsilon = 1 \times 10^{-8}$，权重衰减为 0.1，梯度裁剪阈值为 1.0。整个过程中采用余弦学习率调度器，预热比例为 0.05，并根据开发集损失进行模型选择。

GRPO从相应的SFT检查点初始化，并采用相同的优化、调度器和基础设施设置，但由于内存需求增加，每设备批次大小降至2。GRPO专属超参数包括KL散度系数 $\beta = 0.04$、PPO裁剪范围0.2、价值损失系数0.1和GAE参数 $\lambda = 0.95$。在策略优化期间，我们对每个输入采样16个生成结果，生成长度上限为4096个词元，温度设为1.0，top-p 采样设为0.85。所有实验均使用固定随机种子42运行。

5.5 评估设置与指标

我们遵循 [24] 的评估协议。对于分类，我们报告准确率 *(准确率)*、加权F1 *(加权 F1)* 和宏平均F1 *(宏平均 F1)*。

[4]分类：类别标签；反馈生成：细粒度标签；解释：解释。

解释质量使用 BERTScore *(布里尔分数)* [59] 和 METEOR*(MET)* [6] 进行评估。

6 结果与分析

6.1 与最先进水平的比较

表3展示了定量比较结果，表明我们提出的方法取得了最高的准确率，同时解释质量也具有竞争力[5]。

具体而言，它在分类上达到了81.2%的准确率、0.81的加权F1和0.79的宏平均F1。此外，我们的模型取得了0.78的BERTScore和0.52的METEOR分数，这些结果与最强基线（即 [24]）相当或更高。这些结果凸显了所提出的基于思维的多模态大语言模型以及基于*GRPO*的优化策略 在分类和解释生成任务上的有效性。

模型	Acc.	加权 F1	宏平均 F1	布里尔分数	MET
[22]	69.47±2.06	–	–	–	–
[10]	72.98±1.09	–	–	–	–
[52]	76.4	–	–	–	–
[57]	77.2	–	–	–	–
[8]	77.7	–	–	–	–
[35]	78.8	–	–	–	–
[24]	79.9	0.80	0.79	0.78	0.49
所提方法	**81.2**	**0.81**	**0.79**	**0.78**	**0.52**

表3：与最先进水平及我们结果的比较。

6.1.1 零样本与思维链推理。 表4报告了第5.2节所述实验的结果。该表使用分类和解释生成指标，对开放源和闭源思维型预训练多模态大语言模型进行了基准测试。结果表明，没有任何单一模型能在所有指标上始终取得最高性能。具体而言，L4-Scout的准确率和BERTScore最高；采用思维链提示的商业模型GPT-4.1取得了最佳F1分数，而Qwen的METEOR分数表现最佳。值得注意的是，尽管处于最先进水平，非商业思维型模型和基于思维链的方法都未能在各项评估标准上全面占优。

思维链提示的影响。在所有评估模型中，思维链提示对分类指标和解释指标呈现出系统性且不对称的影响。一方面，思维链提升了大多数模型的分类性能，使加权F1和宏平均F1均稳定提高（Qwen-T最高提升 +4.5个百分点）。更细致的错误分析表明，这些改进主要源于假阴性的减少，说明与直接提示相比，思维链有助于模型更好地识别那些容易被遗漏的细微或隐含仇恨内容。另一方面，以BERTScore衡量时，思维链会持续降低解释质量，而在大多数情况下，METEOR也有所下降。这种下降在所有架构中均有体现，包括开源模型和商业模型，BERTScore降幅介于 −1.6 和 −12.5之间。这一模式表明，尽管思维链增强了模型在决策中的内部推理能力，

[5]所有基线结果均按原始论文所述报告，且未重新实现或重新评估。

模型	准确率	加权 F1	宏平均 F1	BS	Met.
Llama +思维链	60.0	0.606	0.592	0.600	0.165
Llama	64.2	0.599	0.547	0.660	0.234
Δ	-4.2	+0.007	+0.045	-0.060	-0.069
Qwen-I +思维链	65.3	0.658	0.652	0.646	**0.263**
Qwen-I	63.0	0.630	0.630	0.662	0.216
Δ	+2.3	+0.028	+0.022	-0.016	+0.047
Qwen-T +思维链	70.2	0.706	0.694	0.581	0.202
Qwen-T	65.6	0.661	0.652	0.659	0.215
Δ	+4.6	+0.045	+0.042	-0.078	-0.013
Gemma +思维链	65.5	0.659	0.655	0.621	0.135
Gemma	65.2	0.654	0.652	0.664	0.215
Δ	+0.3	+0.005	+0.003	-0.043	-0.080
L4-Scout +思维链	75.4	0.687	0.545	0.596	0.111
L4-Scout	**77.0**	0.710	0.580	**0.653**	0.252
Δ	-1.6	-0.023	-0.035	-0.057	-0.141
GPT-4.1 +思维链	74.1	**0.739**	**0.719**	0.470	0.072
GPT-4.1	71.3	0.715	0.700	0.595	0.159
Δ	+2.8	+0.024	+0.019	-0.125	-0.087

表 4：预训练多模态大语言模型的性能。Llama = Llama- 3.2-11B；L4-Scout = Llama-4-Scout-17B；Qwen-I/T = Qwen3- VL-8B（指令/思考）；Gemma = Gemma-3-12B。Δ 表示（思维链 – 无思维链）。准确率：准确率；W-F1：加权 F1；M-F1：宏平均 F1；BS：BERTScore；度量指标：METEOR。

但它会使输出分布偏离作为参考的简洁标注式解释。

因模型而异的思维链行为。思维链效应的幅度和方向在不同模型家族之间差异显著。Qwen-T 等推理导向型模型从思维链中获益最多，准确率提升幅度最大，这表明思维链提示与其预训练目标之间具有良好的契合度。相比之下，基于Llama的模型准确率提升较弱，甚至有所下降，说明显式思维链可能会干扰其隐式推理机制。有趣的是，GPT-4.1 等大型前沿模型采用思维链后分类性能有所提升，但解释质量下降幅度最大，这凸显了决策准确率与解释校准之间显著的权衡。

模型选择。 这些观察凸显了进一步研究的必要性：如何有效利用开源思考型多模态大语言模型，使其能够在现实的计算预算下进行训练，并支持可复现的后训练实验。综合整体性能与实施便捷性（可在常见 GPU 上训练），我们选择 Qwen 模型作为后续实验的基础多模态大语言模型。

6.1.2 强化学习后训练分析。 表 5 总结了监督微调预热和监督GRPO对分类表现及解释质量的影响。仅监督微调预热就已能持续优于冷启动（不进行预热）的 GRPO。具体而言，**监督微调预热（分类 + 反馈生成 + 解释）**的准确率达到 78.1%，宏平均 F1 为 0.77，优于经训练的变体

模型	准确率	加权 F1	宏平均 F1	布里尔分数	度量指标
监督微调预热					
监督微调（分类+解释）	77.0	0.77	0.75	0.77	0.48
监督微调（分类+反馈生成+解释）	78.1	0.78	0.77	0.77	0.48
监督微调（分类+反馈生成+解释，~~CoTD~~）	79.2	0.79	0.78	0.78	0.50
强化学习 / GRPO					
GRPO（冷启动）	76.8	0.77	0.75	0.73	0.47
监督微调-分类+解释 → GRPO	80.4	0.80	0.78	0.76	0.50
监督微调-分类+反馈生成+解释 → GRPO	**81.1**	**0.81**	0.79	0.77	**0.52**
监督微调-分类+反馈生成+解释-CoTD →GRPO	**81.2**	**0.81**	**0.79**	**0.78**	**0.52**

表5：监督微调预热和GRPO后训练对仇恨模因。分类 =分类标签；反馈生成 =细粒度标签；解释 = 解释；CoTD = 思维链蒸馏-tion

未采用细粒度监督（准确率为 77.0%，宏平均 F1 为 0.75）。这一结果表明，即使缺乏明确的推理监督，纳入细粒度标签也能提升决策一致性。

在监督微调期间加入蒸馏思维链监督可带来进一步提升。**监督微调预热(分类 + 反馈生成 +解释,CoTD)**取得了最强的预热性能，准确率达到 79.2%，宏平均 F1 达到 0.78，同时也提升了解释质量：BERTScore 从 0.77 提升至 0.78，METEOR 从 0.48 提升至 0.50。这些提升表明，监督微调期间的显式推理监督主要提升了解释忠实度，同时也增强了分类性能。

与监督GRPO结合时，所有以监督微调初始化的模型都显著优于此前仅监督微调的模型，以及冷启动（无预热）的 GRPO 训练模型。这一结果表明，未经适当预热或初始化的 *GRPO* 收效有限：冷启动变体的准确率仅为 76.8%，宏平均 F1 为 0.75，解释质量也明显较弱（BERTScore 为 0.73），凸显了监督微调预热的重要性。相比之下，从**监督微调预热**初始化 GRPO 可持续提升性能：监督微调-分类+解释 → *GRPO* 和监督微调-分类+反馈生成+解释 → *GRPO* 均提升了准确率（分别从 77.0% 提升至 80.4%，以及从 78.1% 提升至 81.1%）、宏平均 F1（分别从 0.75 提升至 0.78，以及从 0.77 提升至 0.79）和 METEOR 分数（分别从 0.48 提升至 0.50，以及从 0.48 提升至 0.52）。

接下来，我们分析蒸馏思维链监督对 GRPO 的影响。与我们对仅监督微调模型的观察一致——使用思维链监督时，准确率最高（79.2%）——思维链在基于 GRPO 微调的模型中，也在所有指标上带来了最强且最稳定的表现。具体而言，从经过思维链监督训练的检查点初始化的模型，在 *GRPO* 优化后始终表现最佳，这凸显了思维链蒸馏与强化学习在提升分类和解释质量方面的互补作用。

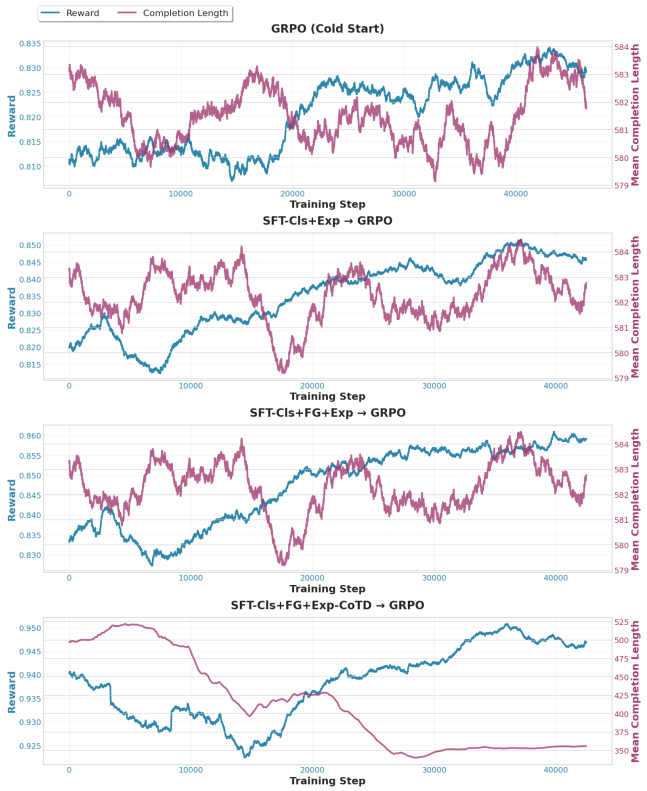


图2：不同初始化方案下的GRPO训练动态，展示平均组奖励（左轴）和平均生成长度（右轴）。采用窗口大小为1000的移动平均平滑处理。

6.2 不同初始化方案下的GRPO训练动态

图 2 比较了不同初始化方案下的 GRPO 奖励变化轨迹和平均生成长度（长度是每个样本生成词元数的平均值）。表 5 报告了相应的最终任务指标。总体而言，GRPO 能显著受益于监督微调预热：预热运行的奖励增长更平稳，最终准确率和解释质量也明显优于冷启动 GRPO。

冷启动GRPO在缺乏先前对齐时表现不佳。直接从预训练骨干网络初始化时，GRPO的奖励曲线噪声更大，收益有限，且生成长度基本稳定。这表明，仅靠强化学习难以同时学会所需的输出结构和任务特定决策边界，这也体现在表5中最低的BS/MET分数上。

监督微调预热让GRPO专注于改进行为，而非学习格式。从SFT（Cls+Exp）初始化GRPO，相较于仅监督微调取得了显著提升（例如，77.0→80.4准确率）。这表明，一旦模型与目标模式及监督分布对齐，GRPO便可专注于提升标签正确性和推理依据质量。

细粒度监督提高了性能上限。使用受保护类别和攻击类型标签进行预热（分类+反馈生成+解释），经过 GRPO 后效果更佳（准确率 81.1，MET 0.52），这表明细粒度监督能够提升决策一致性，并且相较于仅使用二元标签，可以支持更好的解释。

CoTD预热提升了指标，但也揭示了思维链坍缩。CoTD初始化运行在GRPO后取得了整体最佳表现（准确率为81.2，布里尔分数为0.78，MET为0.52），但图2显示，优化过程中生成长度显著下降。这表明模型可能采用了一种强化学习捷径：通过压缩 <think> 片段来提高奖励，而不是持续丰富推理过程，即奖励驱动的思维链坍缩。在内容审核场景中，较短的生成结果仍可能保留标签奖励，同时降低冗长相关风险。

启示。这些动态表明，有必要设计能够明确控制内部推理预算的奖励机制（例如，添加 $R_{\text{think-len}}$或目标范围约束），以抑制坍缩并防止模型利用长度捷径。

7 结论

我们研究了在同时针对预测和解释质量进行显式优化时，基于思维的多模态大语言模型能够在多大程度上提升仇恨梗图理解能力。为此，我们提出了一种采用GRPO目标函数的基于强化学习的后训练方法，该方法联合奖励正确分类、结构有效的输出以及语义一致的解释。此外，我们还构建了一个扩展数据集，其中包含通过弱监督/伪监督蒸馏得到的思维链推理依据，以提供细粒度推理监督。在仇恨表情包基准上的实验表明，我们的方法提升了分类性能和解释质量，取得了最先进的结果，并展现了强化学习后训练在多模态仇恨分析中的价值。未来工作包括*(i)* 在其他仇恨梗图和多模态毒性基准上评估泛化能力，包括多语言场景；*(ii)* 评估该方法在更多具备思考能力的多模态大语言模型及提示方案间的迁移能力；*(iii)* 将该框架扩展到大规模弱标注/伪标注训练，以利用大量未标注梗图；以及*(iv)* 探索面向未标注数据的自监督GRPO目标函数。除了扩展和迁移，我们的训练动态分析还表明，强化学习领域有几个值得探索的方向：*(v)* 设计分段奖励，明确控制内部推理预算，以减轻奖励驱动的思维坍缩。

8 局限性

我们的结果基于仇恨表情包基准，可能无法推广至其他领域或语言。来自商业多模态大语言模型的弱监督思维链蒸馏可能引入偏见和不忠实的推理。GRPO 增加了计算成本，而对闭源模型的依赖可能影响结果的可复现性。解释质量通过自动指标进行评估，但这些指标只能部分反映人类判断。

致谢

A 的Shahroor和F. Alam获得来自卡塔尔国家研究基金的NPRP资助项目14C-0916-210015支持，该基金隶属于卡塔尔研究、发展与创新委员会（QRDI）。本文所报告的研究结果完全由作者负责。

参考文献

[1] Mohammad Awad AlAfnan. 2025年. 模因在塑造社交媒体政治话语中的作用. 媒体与传播研究 13, 2 (2025年), 1–10. [2] Firoj Alam, Abul Hasnat, Fatema Ahmad, Md. Arid Hasan, 和 Maram Hasanain. 2024年. ArMeme：阿拉伯语模因中的宣传性内容. 载于*2024*年自然语言处理实证方法会议论文集, Yaser Al-Onaizan, Mohit Bansal, 和 Yun-Nung Chen（编）. 计算语言学协会, 迈阿密, 佛罗里达州, 美国, 21071–21090. doi:10.18653/v1/2024. emnlp-main.[3] Amalia Amalia, Amer Sharif, Fikri Haisar, Dani Gunawan, 和 Benny B Nasution. 2018年. 使用光学字符识别（OCR）和朴素贝叶斯算法进行模因观点分类. 载于*2018*年第三届信息学与计算国际会议（*ICIC*）. IEEE, 1–5.[4] Shuai Bai, Yuxuan Cai, Ruizhe Chen, Keqin Chen, Xionghui Chen, Zesen Cheng, Lianghao Deng, Wei Ding, Chang Gao, Chunjiang Ge, Wenbin Ge, Zhifang Guo, Qidong Huang, Jie Huang, Fei Huang, Binyuan Hui, Shutong Jiang, Zhaohai Li, Mingsheng Li, Mei Li, Kaixin Li, Zicheng Lin, Junyang Lin, Xuejing Liu, Jiawei Liu, Chenglong Liu, Yang Liu, Dayiheng Liu, Shixuan Liu, Dunjie Lu, Ruilin Luo, Chenxu Lv, Rui Men, Lingchen Meng, Xuancheng Ren, Xingzhang Ren, Sibo Song, Yuchong Sun, Jun Tang, Jianhong Tu, Jianqiang Wan, Peng Wang, Pengfei Wang, Qiuyue Wang, Yuxuan Wang, Tianbao Xie, Yiheng Xu, Haiyang Xu, Jin Xu, Zhibo Yang, Mingkun Yang, Jianxin Yang, An Yang, Bowen Yu, Fei Zhang, Hang Zhang, Xi Zhang, Bo Zheng, Humen Zhong, Jingren Zhou, Fan Zhou, Jing Zhou, Yuanzhi Zhu, 和 Ke Zhu. 2025年. Qwen3-VL技术报告. *arXiv* 预印本 *arXiv:2511.21631*. https://arxiv.org/abs/2511.21631.[5] Satanjeev Banerjee 和 Alon Lavie. 2005年. METEOR：一种与人类判断相关性更高的机器翻译评价自动指标. 载于机器翻译和/或摘要的内在与外在评价指标*ACL*研讨会论文集. 65–72.[6] Satanjeev Banerjee 和 Alon Lavie. 2005年. METEOR：一种与人类判断相关性更高的机器翻译评价自动指标. 载于机器翻译和/或摘要的内在与外在评价指标*ACL*研讨会论文集, Jade Goldstein, Alon Lavie, Chin-Yew Lin, 和 Clare Voss（编）. 计算语言学协会, 安娜堡, 密歇根州, 65–72. https://aclanthology.org/W05-0909/ [7] Sravani Boinepelli, Manish Shrivastava, 和 Vasudeva Varma. 2020年. Sis@ iiith在SemEval-2020 任务 8上的研究：模因分析中简单文本分类方法概述. 载于第十四届语义评测研讨会论文集. 1190–1194.[8] Giovanni Burbi, Alberto Baldrati, Lorenzo Agnolucci, Marco Bertini, 和 Alberto Del Bimbo. 2023年. 将模因映射为词语以进行多模态仇恨模因分类. 载于*IEEE/CVF*计算机视觉国际会议论文集. 2832–2836.[9] Rui Cao, Ming Shan Hee, Adriel Kuek, Wen-Haw Chong, Roy Ka-Wei Lee, 和 Jing Jiang. 2023年. Pro-cap：利用冻结的视觉语言模型进行仇恨模因检测. 载于第*31*届*ACM*国际多媒体会议论文集. 5244–5252.[10] Rui Cao, Roy Ka-Wei Lee, Wen-Haw Chong, 和 Jing Jiang. 2022年. 面向多模态仇恨模因分类的提示. 载于*2022*年自然语言处理实证方法会议论文集, Yoav Goldberg, Zornitsa Kozareva, 和 Yue Zhang（编）. 计算语言学协会, 阿布扎比, 阿拉伯联合酋长国, 321–332. doi:10.18653/v1/2022.emnlp-main.[11] Rui Cao, Roy Ka-Wei Lee, 和 Jing Jiang. 2024年. 用于少样本仇恨模因检测的模块化网络. 载于*2024*年*ACM*网络会议论文集. 4575–4584.[12] Jiajun Chen 和 Yik-Cheung Tam. 2025年. 面向数学推理的谓词引导生成. 载于*2025*年自然语言处理实证方法会议论文集. 9097–9110.[13] Gheorghe Comanici, Eric Bieber, Mike Schaekermann, Ice Pasupat, Noveen Sachdeva, Inderjit Dhillon, Marcel Blistein, Ori Ram, Dan Zhang, Evan Rosen, 等. 2025年. Gemini 2.5：借助先进推理、多模态、长上下文和下一代智能体能力推动前沿. *arXiv* 预印本 *arXiv:2507.06261*(2025年). [14] DeepSeek-AI, Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, Xiaokang Zhang, Xingkai Yu, Yu Wu, Z. F. Wu, Zhibin Gou, Zhihong Shao, Zhuoshu Li, Ziyi Gao, Aixin Liu, Bing Xue, Bingxuan Wang, Bochao Wu, BeiFeng, Chengda Lu, Chenggang Zhao, Chengqi Deng, Chenyu Zhang, Chong Ruan, Damai Dai, Deli Chen, Dongjie Ji, Erhang Li, Fangyun Lin, Fucong Dai, Fuli Luo, Guangbo Hao, Guanting Chen, Guowei Li, H. Zhang, Han Bao, Hanwei Xu, Haocheng Wang, Honghui Ding, Huajian Xin, Huazuo Gao, Hui Qu, Hui Li, Jianzhong Guo, Jiashi Li, Jiawei Wang, Jingchang Chen, Jingyang Yuan, Junjie Qiu, Junlong Li, J. L. Cai, Jiaqi Ni, Jian Liang, Jin Chen, Kai Dong, Kai Hu, Kaige Gao, Kang Guan, Kexin Huang, Kuai Yu, Lean Wang, Lecong Zhang, Liang Zhao, Litong Wang, Liyue Zhang, Lei Xu, Leyi Xia, Mingchuan Zhang, Minghua Zhang, Minghui Tang, Meng Li, Miaojun Wang, Mingming Li, Ning Tian, Panpan Huang, Peng Zhang, Qiancheng Wang, Qinyu Chen, Qiushi Du, Ruiqi Ge, Ruisong Zhang, Ruizhe Pan, Runji Wang, R. J. Chen, R. L. Jin, Ruyi Chen, Shanghao Lu, Shangyan Zhou, Shanhuang Chen, Shengfeng Ye, Shiyu Wang, Shuiping Yu, Shunfeng Zhou, Shuting Pan, S. S. Li, Shuang Zhou, Shaoqing Wu, Shengfeng Ye, Tao Yun, Tian Pei, Tianyu Sun, T. Wang, Wangding Zeng, Wanjia Zhao, Wen Liu, Wenfeng Liang, Wenjun Gao, Wenqin Yu, Wentao Zhang, W. L. Xiao, Wei An, Xiaodong Liu, Xiaohan Wang, Xiaokang Chen, Xiaotao Nie, Xin Cheng, Xin Liu, Xin Xie, Xingchao Liu, Xinyu Yang, Xinyuan Li, Xuecheng Su, Xuheng Lin, X. Q. Li, Xiangyue Jin, Xiaojin Shen, Xiaosha Chen, Xiaowen Sun, Xiaoxiang Wang, Xinnan Song, Xinyi Zhou, Xianzu Wang, Xinxia Shan, Y. K. Li, Y. Q. Wang, Y. X. Wei, Yang Zhang, Yanhong Xu, Yao Li, Yao Zhao, Yaofeng Sun, Yaohui Wang, Yi Yu, Yichao Zhang, Yifan Shi, Yiliang Xiong, Ying He, Yishi Piao, Yisong Wang, Yixuan Tan, Yiyang Ma, Yiyuan Liu, Yongqiang Guo, Yuan Ou, Yuduan Wang, Yue Gong, Yuheng Zou, Yujia He, Yunfan Xiong, Yuxiang Luo, Yuxiang You, Yuxuan Liu, Yuyang Zhou, Y. X. Zhu, Yanhong Xu, Yanping Huang, Yaohui Li, Yi Zheng, Yuchen Zhu, Yunxian Ma, Ying Tang, Yukun Zha, Yuting Yan, Z. Z. Ren, Zehui Ren, Zhangli Sha, Zhe Fu, Zhean Xu, Zhenda Xie, Zhengyan Zhang, Zhewen Hao, Zhicheng Ma, Zhigang Yan, Zhiyu Wu, Zihui Gu, Zijia Zhu, Zijun Liu, Zilin Li, Ziwei Xie, Ziyang Song, Zizheng Pan, Zhen Huang, Zhipeng Xu, Zhongyu Zhang, and Zhen Zhang. 2025年. DeepSeek-R1：通过强化学习激发大语言模型的推理能力。arXiv:2501.12948 [cs.CL] https://arxiv.org/abs/2501.12948[15] Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Amy Yang, Angela Fan, et al. 2024年。Llama 3模型家族。*arXiv* 预印本 *arXiv:2407.21783* (2024年)。[16] Elisabetta Fersini, Francesca Gasparini, Giulia Rizzi, Aurora Saibene, Berta Chulvi, Paolo Rosso, Alyssa Lees, and Jeffrey Sorensen. 2022年。SemEval-2022任务5：多模态自动厌女识别。载于第*16*届语义评估国际研讨会（*SemEval-2022*）论文集, Guy Emerson, Natalie Schluter, Gabriel Stanovsky, Ritesh Kumar, Alexis Palmer, Nathan Schneider, Siddharth Singh, and Shyam Ratan（编）。计算语言学协会，美国西雅图，533–549. doi：。/v / .semeval-。[17] Biagio Grasso, Valerio La Gatta, Vincenzo Moscato, and Giancarlo Sperlì. 2024年。Kermit：有害梗图检测中的知识赋能模型。*Information Fusion* 106（2024年），102269。[18] MS Hee, A Kumaresan, and RKW Lee. [无日期]。跨越模态：通过少样本上下文学习增强跨模态仇恨言论检测（2024年）。[19] Lawrence R James, Robert G Demaree, and Gerrit Wolf. 1984年。估算组内评分者间信度，并考虑反应偏差。应用心理学杂志69, 1（1984年），85。[20] L. P. Kaelbling, M. L. Littman, and A. W. Moore. 1996年。强化学习：综述。arXiv:cs/9605103 [ ] 9605103 cs.AI https://arxiv.org/abs/cs/ Timo Kaufmann, Paul Weng, Viktor Bengs, and Eyke Hüllermeier. 2024年。基于人类反馈的强化学习综述。（2024年）。[22] Douwe Kiela, Hamed Firooz, Aravind Mohan, Vedanuj Goswami, Amanpreet Singh, Pratik Ringshia, and Davide Testuggine. 2020年。仇恨模因挑战赛：检测多模态梗图中的仇恨言论。神经信息处理系统进展33（2020年），2611–2624。[23] Douwe Kiela, Hamed Firooz, Aravind Mohan, Vedanuj Goswami, Amanpreet Singh, Pratik Ringshia, and Davide Testuggine. 2021年。仇恨模因挑战赛：检测多模态梗图中的仇恨言论。arXiv:2005.04790 [cs.AI] https://arxiv.org/abs/2005.04790[24] Mohamed Bayan Kmainasi, Abul Hasnat, Md Arid Hasan, Ali Ezzat Shahroor, and Firoj Alam. 2025年。MemeIntel：宣传性和仇恨梗图的可解释检测。载于*2025*年自然语言处理实证方法会议论文集，Christos Christodoulopoulos, Tanmoy Chakraborty, Carolyn Rose, and Violet Peng（编）。计算语言学协会，中国苏州，30251–30267. doi：。/v / .emnlp-main。[25] Gitanjali Kumari, Kirtan Jain, and Asif Ekbal. 2024年。M hop-cot：通过多模态多跳思维链识别厌女梗图。*arXiv* 预印本 *arXiv:2410.09220* （2024年）。[26] Ehsan Latif, Yifan Zhou, Shuchen Guo, Yizhu Gao, Lehong Shi, Matthew Nyaaba, Arne Bewerdorff, Xiantong Yang, Xiaoming Zhai, et al. 2025年。OpenAI O1与人类在高阶认知方面的比较评估。*Scientific Reports* （2025年）。doi:10.1038/s41598-025-33629-9[27] Mingxin Li, Yanzhao Zhang, Dingkun Long, Keqin Chen, Sibo Song, Shuai Bai, Zhibo Yang, Pengjun Xie, An Yang, Dayiheng Liu, et al. 2026年。Qwen3-VL-Embedding和Qwen3-VL-Reranker：用于最先进多模态检索与排序的统一框架。*arXiv* 预印本 *arXiv:2601.04720* （2026年）。[28] Hongzhan Lin, Ziyang Luo, Wei Gao, Jing Ma, Bo Wang, and Ruichao Yang. 2024年。通过大语言模型之间的多模态辩论实现可解释的有害梗图检测。载于*2024*年*ACM*网络会议论文集。2359–2370。[29] Hongzhan Lin, Ziyang Luo, Jing Ma, and Long Chen. 2023年。表象之下：从大语言模型蒸馏多模态推理以揭示有害梗图。载于计算语言学协会：*2023*年*EMNLP*论文集，Houda Bouamor, Juan Pino, and Kalika Bali（编）。计算语言学协会，新加坡，9114–9128。doi:10.18653/v1/2023.findings-emnlp.611

[30] Ziyan Liu, Chunxiao Fan, Haoran Lou, Yuexin Wu, and Kaiwei Deng. 2025年。MIND：用于零样本有害表情包检测的多智能体框架。*arXiv* 预印本 *arXiv:2507.06908*（2025年）。[31] Junyu Lu, Bo Xu, Xiaokun Zhang, Haohao Zhu, Kaichun Wang, Liang Yang, and Hongfei Lin. 2025年。有推理依据就够了吗？重新思考多模态仇恨模因检测中的知识增强。载于第*48*届国际 *ACM SIGIR* 信息检索研究与发展大会论文集。559–569。[32] Pan Lu, Hritik Bansal, Tony Xia, et al. 2024年。MathVista：评估基础模型在视觉上下文中的数学推理能力。载于国际学习表征会议（*ICLR*）。https://arxiv.org/abs/ 2310 02255 。Lambert [33] Mathias, Shaoliang Nie, Aida Mostafazadeh Davani, Douwe Kiela, Vin-odkumar Prabhakaran, Bertie Vidgen, and Zeerak Talat. 2021年。WOAH 5共享任务：细粒度仇恨表情包检测的研究结果。载于第*5*届在线虐待与伤害研讨会论文集（*WOAH 2021*）。201–206。[34] Lambert Mathias, Shaoliang Nie, Aida Mostafazadeh Davani, Douwe Kiela, Vin-odkumar Prabhakaran, Bertie Vidgen, and Zeerak Waseem. 2021年。WOAH 5共享任务：细粒度仇恨表情包检测的研究结果。载于第*5*届在线虐待与伤害研讨会论文集（*WOAH 2021*），Aida Mostafazadeh Davani, Douwe Kiela, Mathias Lambert, Bertie Vidgen, Vinodku-mar Prabhakaran, and Zeerak Waseem（编）。计算语言学协会，线上， 201 206 10 18653 1 2021 1 21 –。doi：。/v / .woah-。[35] Jingbiao Mei, Jinghong Chen, Weizhe Lin, Bill Byrne, and Marcus Tomalin. 2024年。通过检索引导的对比学习改进仇恨模因检测。载于计算语言学协会第*62*届年会论文集（第*1*卷：长篇论文）。5333–5347。[36] Jingbiao Mei, Mingsheng Sun, Jinghong Chen, Pengda Qin, Yuhong Li, Da Chen, and Bill Byrne. 2025年。ExPO-HM：学习先解释后检测，以检测仇恨模因。*arXiv* 预印本 *arXiv:2510.08630*（2025年）。[37] Mihaela-Georgiana Mihăilescu. 2024年。千万别惹"表情包创作者"：表情包创作者如何重新定义当代政治。社交媒体 + 10, 与社会（），。4 2024 20563051241296256。[38] R OpenAI. 2023年。GPT-4技术报告。*arXiv*（2023年），2303–08774。[39] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. 2022年。训练语言模型，使其能够遵循指令并利用人类反馈。神经信息处理系统进展35（2022年），27730–27744。[40] Delfina S Martinez Pandiani, Erik Tjong Kim Sang, and Davide Ceolin. 2025年。"有毒"模因：关于表情包毒性检测与解释的计算视角综述。*Online Social NetworksandMedia* 47（2025年），100317。[41] Shraman Pramanick, Shivam Sharma, Dimitar Dimitrov, Md Shad Akhtar, Preslav Nakov, and Tanmoy Chakraborty. 2021年。MOMENTA：用于检测有害梗图及其目标的多模态框架。*arXiv* 预印本 *arXiv:2109.05184*（2021年）。[42] Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. 2023年。直接偏好优化：你的语言模型其实暗中充当奖励模型。神经信息处理系统进展36（2023年），53728–53741。[43] Leonardo Ranaldi and Giulia Pucci. 2025年。通过自训练实现多语言推理。载于*2025*年计算语言学协会美洲分会会议：人类语言技术论文集（第*1*卷：长篇论文）。11566–11582。[44] Ursula Kristin Schmid. 2025年。社交媒体上的幽默仇恨言论：对用户如何感知和处理仇恨模因的混合方法研究。*New Media & Society* 27, 3（2025年），1588–1606。[45] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017年。近端策略优化算法。*arXiv* 预印本 *arXiv:1707.06347*（2017）。[46] Zhihong Shao、Peiyi Wang、Qihao Zhu、Runxin Xu、Junxiao Song、Xiao Bi、Haowei Zhang、Mingchuan Zhang、Y. K. Li、Y. Wu 和 Daya Guo。2024。DeepSeek-Math：突破开放语言模型数学推理能力的极限。arXiv:2402.03300 [cs.CL] https://arxiv.org/abs/2402.03300[47] Shivam Sharma、Firoj Alam、Md Shad Akhtar、Dimitar Dimitrov、Giovanni Da San Martino、Hamed Firooz、Alon Halevy、Fabrizio Silvestri、Preslav Nakov 和 Tanmoy Chakraborty。2022。检测和理解有害梗图：一项综述。*arXiv* 预印本 *arXiv:2205.04274*（2022）。[48] Ingroj Shrestha 和 Jonathan Rusert。2020。NLP_UIOWA 在 SemEval-2020 任务 8 中的系统：你并非唯一一个被知识诅咒的人——用于模因情感分析的多分支模型。载于第十四届语义评测研讨会论文集。891–900。[49] Li Sun、Chaitanya Ahuja、Peng Chen、Matt D'Zmura、Kayhan Batmanghelich 和 Philip Bontrager。2025。多模态大语言模型是有效的视觉学习器。载于*2025*年 *IEEE/CVF* 冬季计算机视觉应用会议（*WACV*）。IEEE，8617–8626。[50] Gemma 团队、Aishwarya Kamath、Johan Ferret、Shreya Pathak、Nino Vieillard、Ramona Merhej、Sarah Perrin、Tatiana Matejovicova、Alexandre Ramé、Morgane Rivière 等。2025。Gemma 3 技术报告。*arXiv* 预印本 *arXiv:2503.19786*（2025）。[51] Lai Wei、Yuting Li、Chen Wang、Yue Wang、Linghe Kong、Weiran Huang 和 Lichao Sun。2025。先进行监督微调，再进行强化学习，最后进行无监督后训练：通过无监督后训练持续改进多模态大语言模型推理。arXiv:2505.22453 [cs.CL] https://arxiv.org/abs/ 2505 22453 。Fan Wu、Guolian Chen、Junkuo Cao、[52] Yuhan Yan 和 Zhongneng Li。2024。基于迁移学习和交叉掩码机制的多模态仇恨模因分类。*Electronics* 13，14（2024），2780。[53] Jiannan Wu、Muyan Zhong、Sen Xing、Zeqiang Lai、Zhaoyang Liu、Zhe Chen、Wenhai Wang、Xizhou Zhu、Lewei Lu、Tong Lu 等。2024。VisionLLM v2：面向数百种视觉语言任务的端到端通用多模态大语言模型。神经信息处理系统进展37（2024），69925–69975。[54] Xiaobao Wu。2025。仰望星辰：奖励模型、学习策略与从奖励中学习综述。*arXiv* 预印本 *arXiv:2505.02686*（2025）。[55] Yanzhi Xu、Yueying Hua、Shichen Li 和 Zhongqing Wang。2024。探索思维链在多模态隐喻检测中的应用。载于计算语言学协会第*62*届年会论文集（第一卷：长篇论文）。91–101。[56] An Yang、Anfeng Li、Baosong Yang、Beichen Zhang、Binyuan Hui、Bo Zheng、Bowen Yu、Chang Gao、Chengen Huang、Chenxu Lv 等。2025。Qwen3 技术报告。*arXiv* 预印本 *arXiv:2505.09388*（2025）。[57] Chuanpeng Yang、Yaxin Liu、Fuqing Zhu、Jizhong Han 和 Songlin Hu。2024。用于仇恨模因检测的不确定性引导模态再平衡。载于计算语言学协会第*62*届年会论文集（第一卷：长篇论文）。4361–4371。[58] Xiang Yue、Yuansheng Ni、Kai Zhang 等。2024。MMMU：面向专家级通用人工智能的大规模多学科多模态理解与推理基准。载于*IEEE/CVF* 计算机视觉与模式识别会议（*CVPR*）论文集。https://arxiv.org/abs/ 2311 16502 。Tianyi Zhang、Varsha Kishore、[59] Felix Wu、Kilian Q. Weinberger 和 Yoav Artzi。2020。BERTScore：使用 BERT 评估文本生成。arXiv: 1904 09675 [ ] 1904 09675 。[60] cs.CL https://arxiv.org/abs/ 。Xunyu Zhu、Jian Li、Yong Liu、Can Ma 和 Weiping Wang。2024。将数学推理能力蒸馏至小型语言模型。*NeuralNetworks* 179（2024），106594。