arXiv:2601.01392v1 [cs.SD] 4 Jan 2026

SAFE-QAQ：通过强化学习实现端到端慢思考音频文本欺诈检测

PeidongWang[1]* ZhimingMa[1], 2* Xin Dai[1]* Yongkang Liu[1] Shi Feng[1]†Xiaocui Yang[1] Wenxing Hu[3] ZhihaoWang[1] Mingjun Pan[2] Li Yuan[4] DalingWang[1]中国，东北大学[2]中国移动互联网有限公司[3]中国上海，上海电力大学[4]中国深圳，北京大学 pdongwang@163.com, mazhiming312@outlook.com daix1@mails.neu.edu.cn, fengshi@cse.neu.edu.cn

摘要

现有欺诈检测方法主要依赖转录文本，容易受到自动语音识别错误的影响，并会遗漏声音语调和环境语境等关键声学线索。这限制了它们应对复杂欺骗策略的效果。为应对这些挑战，我们首先提出**SAFE-QAQ**，这是一种端到端的综合音频慢思考欺诈检测框架。首先，SAFE-QAQ 框架消除了语音转录错误对检测性能的影响。其次，我们提出基于规则的慢思考奖励机制，通过层次化推理过程准确捕捉细粒度音频细节，系统性地引导系统识别欺诈指示模式。此外，我们的框架在实时通话期间引入动态风险评估，实现欺诈的早期检测与预防。在 TeleAntiFraud-Bench 上进行的实验表明，SAFE-QAQ 在准确性、推理效率和实时处理能力等多个关键维度上均显著优于现有方法。目前，SAFE-QAQ 已投入使用，每天分析超过 70,000 通电话，能够有效自动化复杂的欺诈检测，减少人工工作量和财务损失。代码：https://anonymous.4open.science/r/SAFE-QAQ。

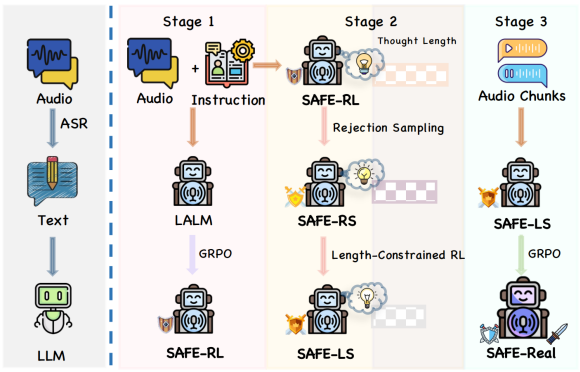

图 1：（a）以往方法与（b）我们提出的方法的比较：通过强化学习（RL）实现端到端通话欺诈检测。我们的方法分三个阶段训练大型音频语言模型（LALMs）：（i）通过强化学习开发慢思考推理；（ii）利用拒绝采样微调和长度约束强化学习优化思考长度；（iii）通过音频块训练实现实时检测。

1 引言

随着移动通信技术的快速发展，电信诈骗问题日益严重，已成为全球性的社会挑战。如图1（a）所示，大多数现有的欺诈检测方法（Chang 等人，2024年；Ko-rkanti，2024年；Yang 等人，2025年；Shen 等人，2025b；Hu 等人，2024年）通过利用转录文本取得了显著进展，文本，以及大语言模型（LLMs）的强大表征能力（Dubey 等人，2024年；Achiam 等人，2023年；Guo 等人，2025年）。然而，由于无法及时更新信息和策略，这些方法往往会导致实时欺诈检测系统作出误判。为解决这一问题，研究人员提出采用检索增强生成（RAG），以获取最新信息和策略，从而无需重新训练（Singh 等人，2025年）。尽管做出了这些努力，欺诈检测系统在现实场景中的应用仍然面临挑战，远未取得全面成功，主要有以下三个原因。第一，自动语音识别（ASR）系统将语音转录为文本时可能引入噪声错误，导致模型性能下降（Ma 等人，2025年；Hu 等人，2024年；Chakraborty 等人，2024年）。第二，基于文本的欺诈检测系统无法捕捉语音传达的

* 这些作者对本研究做出了同等贡献。
† 通讯作者。

细粒度信息，例如声音语调、情绪压力和环境声学特征，而这些信息对于有效的欺诈检测至关重要。第三，现代欺诈者会采用复杂的分层欺骗策略——结合操纵的语音模式（如语音伪造或合成音频）、伪造的背景声音（如伪造的呼叫中心噪声）以及心理胁迫话术——需要通过迭代推理才能揭示其真相（Ma 等人，2025年）。然而，当前基于文本的流水线方法缺乏深度推理机制，无法有效应对这些复杂情况。这些局限性凸显了仅依赖文本的方法在应对现代电信欺诈和复杂欺骗策略方面的不足。

为应对这些挑战，我们提出 SAFE-QAQ（**S**low-thinking **A**udio-text **F**raud d**E**tectionusing **Q**wen **A**udio with **Q**uestion），这是一个端到端的综合框架，用于基于音频的慢思考欺诈检测。基于大型音频语言模型（LALMs）的最新进展（Chu 等人，2024年；Huang 等人，2025年；Hurst 等人，2024年；Zeng 等人，2024年），SAFE-QAQ 构建了一个完整的端到端流程，可直接处理原始音频信号，在保留关键多模态特征的同时融入三项关键创新（如图1（b）所示）：

- SAFE-QAQ 通过基于规则的奖励培养慢思考推理，使系统能够系统地分析细粒度细节，而这些细节往往被采用分层欺骗策略的欺诈者所掩盖。
- 我们进一步通过拒绝采样微调（生成 *SAFE-RS*）和长度约束强化学习（生成 *SAFE-LS*），将推理链长度缩短 48.87%，从而优化推理效率，确保推理简洁而准确。
- SAFE-QAQ 通过结构化提示和阶段识别奖励，动态评估信息充分性，实现实时检测；最终形成的模型（*SAFE-Real*）能够在实时通话中及时进行干预。

通过消除对易出错的自动语音识别转录的依赖，并将慢思考推理与经强化学习优化的多模态处理相结合，SAFE-QAQ 构建了一个完整的端到端框架，同时实现了高准确性和实用效率。其可行性已通过成功部署在一条每天处理超过 70,000 通电话的生产流水线上得到验证；该框架通过及时的自动干预，有效减轻了人工审核负担并防止了财务损失。

2 相关工作

2.1 基于大语言模型的电信诈骗检测

大语言模型的最新进展为电信欺诈检测带来了希望，例如用于实时通话分析的检索增强生成（RAG）（Singh 等人，2025年）以及基于意图的预警系统（Shen 等人，2025b）。然而，这些方法仅依赖转录文本，忽略了能够表明欺诈行为的重要音频特征（如语气、情绪）（Chang 等人，2024年）。TeleAntiFraud-28k 等多模态基准（Ma 等人，2025年）虽弥补了这一不足，但其监督微调（SFT）方法未能充分利用现代大语言模型的推理能力。当前系统还存在效率低下的问题，需要通过多阶段流程完成语音转录和欺诈检测（Yang 等人，2025年）。SAFE-QAQ 通过强化学习（RL）端到端处理原始音频，克服了这些局限，在保留多模态线索的同时消除了中间步骤。这种方法能够实现更快速、更准确的欺诈检测，并适应现实场景的动态变化。

2.2 大型音频语言模型（LALM）

Qwen2-Audio、GLM-4-Voice、GPT-4o 和 Step-Audio 等大型音频语言模型在语音理解方面表现出色，能够实时捕捉语气、情绪和意图（Chu 等人，2024年；Zeng 等人，2024年；Hurst 等人，2024年；Huang 等人，2025年），但它们在欺诈检测中的应用仍然有限。通用大型音频语言模型往往无法识别剧本式欺骗，即语音表达与剧本中刻意保持的平静语气之间存在细微矛盾。**SAFE-QAQ**通过将音频语言建模与面向电信欺诈检测的领域专用强化学习优化相结合，弥合了这一差距，从而实现具备上下文感知能力、对风险敏感的推理。

2.3 面向慢思考的强化学习

强化学习（RL）的最新进展使大语言模型能够发展出慢思考能力，模仿人类式的系统 2 推理（Kahneman，2011）。诸如 Ope-

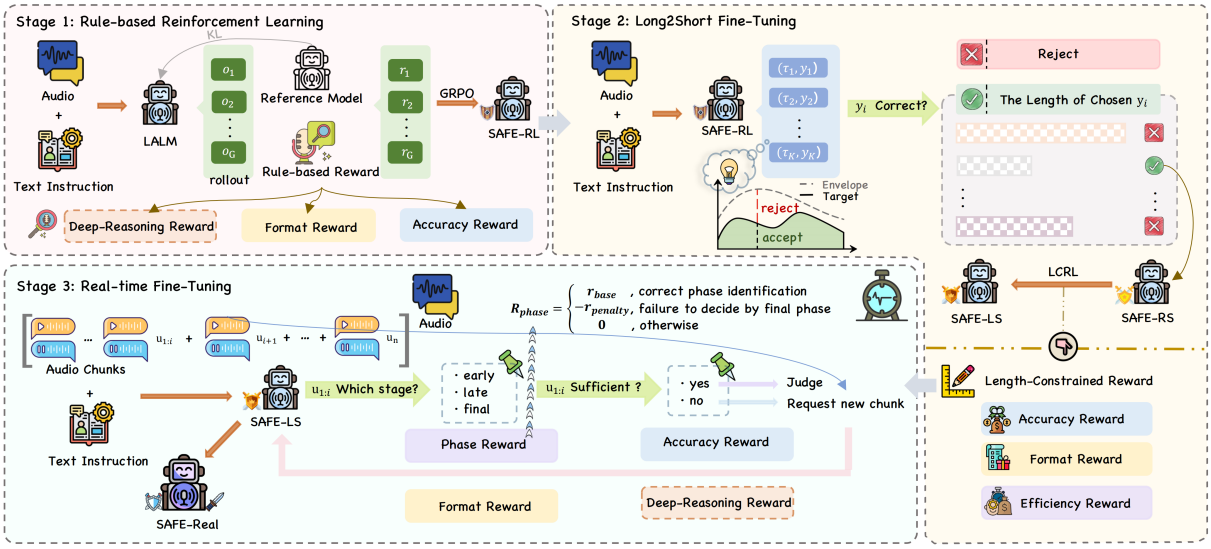

图 2：我们的方法概览。从一个大型音频语言模型（LALM）出发，我们：(i) 应用基于规则的强化学习，获得具备慢思考能力的 SAFE-RL；(ii) 使用拒绝采样（SAFE-RS）和长度约束强化学习（SAFE-LS）对其进行优化，以提升推理效率；以及 (iii) 在音频块上执行实时微调，得到 SAFE-Real。

nAI 的 o1/o3（Jaech 等人，2024年；OpenAI，2025年）、DeepSeek R1（Guo 等人，2025年）和 Satori（Shen 等人，2025年 a）在需要逐步分析的任务中表现出显著提升，例如数学（Zhu 等人，2022年；Lu 等人）、逻辑（Jin 等人，2024年）和多模态推理（Xu 等人，2024年；Thawakar 等人，2025年）。这些模型利用蒙特卡洛树搜索（MCTS）（Swiechowski 等人，2023年）和奖励引导的微调（Trung 等人，2024年）等技术生成更长的推理链，从而提升解决复杂问题的能力。然而，当前基于强化学习的方法主要聚焦于基于文本的推理，对音频与文本集成等多模态领域的探索仍不充分。过度思考（Chen 等人，2024年）以及在动态场景中效率低下（Qi 等人，2024年）等挑战，凸显了制定更有针对性的解决方案的必要性。

3 方法

图 2展示了我们方法的三阶段框架。在阶段 1 中，我们使用基于规则的强化学习，训练具备慢思考能力的模型。在阶段 2 中，我们通过长转短微调对其进行优化，以缩短推理过程并缓解过度思考。最后，在阶段 3 中，我们应用实时微调，优化模型以实现高效的实时欺诈检测。

3.1 问题定义

该任务涉及基于音频分析的三个分类目标：**场景分类**，**欺诈检测**，以及**欺诈类型分类**。给定由原始音频 u和文本指令 t组成的输入对 (u, t)，模型 π 生成输出$o = (\tau, y)$，其中 τ 是逐步推理过程，y 包含分类依据和最终结果。目标是开发能够准确执行这些任务，同时提供可解释推理的 π。

3.2 基于规则的强化学习

如图 2所示，阶段 1 采用基于规则的强化学习进行数据高效的自我演化，生成慢思考模型（SAFE-RL）。该模型能够分析音频文本线索，检测出仅依赖文本的方法难以发现的细微欺诈模式。

组相对策略优化。 与近端策略优化（PPO）等传统的演员-评论家算法（Schulman 等人，2017年）不同，我们使用组相对策略优化（GRPO）来优化模型，无需使用参数复杂度与策略模型相当的评论家模型π_θ。相反，GRPO 基于组内评分来估计每个回答的相对优势。具体而言，对于每个音频文本指令对（u，t）$\sim P(U, T)$，策略模型 π_θ 会采样多个推理过程及其对应的回答。第 i个样本的输出表示为 $o_i = (\tau_i, y_i)$。每个回答 y_i 都会由基于规则的奖励模型评估，以计算奖励值 $r_i = R(y_i)$。根据这些奖励值计算得出的组内相对优势 $A_i = \frac{r_i - mean(\{r1,r2,...,r_G\})}{std(\{r1,r2,...,r_G\})}$，

随后用于通过目标函数 $J_{GRPO}(\theta)$ 优化模型：

$$J_{GRPO}(\theta) = \mathbb{E}_{(u,t)\sim P(U,T),\{o_i\}_{i=1}^{G}\sim\pi_{\theta_{old}}(O|u,t)} \frac{1}{G} \sum_{i=1}^{G}(min(\rho_i A_i, clip_\epsilon(\rho_i)A_i) - \beta\mathbb{D}_{KL}(\pi_\theta \parallel \pi_{ref})) \quad (1)$$

重要性采样因子 $\rho_i(\theta)$ 定义为当前策略 π_θ 与采样策略 $\pi_{\theta_{\text{old}}}$ 的比值。裁剪函数 $clip_\epsilon(\rho_i)$ 将 ρ_i 限制在区间 $[1-\epsilon, 1+\epsilon]$ 内，从而确保保守策略更新。超参数 β 控制 KL 散度 $\mathbb{D}_{KL}$ 的强度，其无偏估计方式如下：

$$\mathbb{D}_{KL}(\pi_\theta||\pi_{ref}) = \frac{\pi_{ref}(o_i|u,t)}{\pi_\theta(o_i|u,t)} - \log\frac{\pi_{ref}(o_i|u,t)}{\pi_\theta(o_i|u,t)} - 1 \quad (2)$$

奖励建模。训练基于结果或基于过程的神经奖励模型既复杂又耗费资源。将通用神经奖励模型迁移到特定领域，需要大量数据和计算资源。相比之下，只需为答案设计验证规则，基于规则的奖励模型便能有效地对奖励进行建模。

• **准确性奖励**（R_{acc}）：验证从 <answer> 标签中提取的最终答案 y_i：

$$R_{acc} = \mathbb{I}(y_i \text{ is correct}) \quad (3)$$

• **格式奖励**（R_{fmt}）：通过 <think> 和 <answer> 标签确保结构：

$$R_{fmt} = \mathbb{I}(\text{format is satisfied}) \quad (4)$$

•**深度推理奖励**（R_{depth}）：使用长度敏感奖励促进更深入的推理。对数归一化可提高对较短推理链的敏感度：

$$R_{depth} = \min\left(\frac{\ln(|\tau|+1)}{\ln(L_{max}+1)}, 1\right)\cdot R^{max} \quad (5)$$

其中 $|\tau|$ 表示推理步数，L_{max} 表示令牌上限，R^{max} 表示奖励上限。

总奖励 R_{total} 的计算方式如下：

$$R_{total} = \alpha R_{acc} + \beta R_{fmt} + \gamma R_{depth} \quad (6)$$

我们设置权重 $\alpha = 5$ 和 $\beta = 1$，以优先保证准确性和格式，并定义 $\gamma = \mathbb{I}_{\text{non-SFT}}$，以鼓励非 SFT 模型进行深入推理。此配置在可靠的欺诈检测与结构化、全面的分析之间实现了平衡。

3.3 长转短微调

如图 2所示，阶段 2 通过 **长转短微调**优化效率。该阶段结合拒绝采样（SAFE-RS）和长度约束强化学习（SAFE-LS），在不牺牲精确性的情况下压缩推理链。

拒绝采样微调。从提议分布$\pi_\theta(\cdot|u,\ t)$中采样 K 个候选项 $\{o_i = (\tau_i,\ y_i)\}_{i=1}^{K}$，我们定义目标分布 π^*，以优先保证正确性和简洁性：

$$\pi^*(o|u,t) \propto \mathbb{I}(y \text{ is correct}) \cdot (1+|\tau|)^{-1} \quad (7)$$

通过最大化以下指标来选出最优响应 o^*：

$$o^* = \arg\max_{i\in[K]} \pi^*(o_i|u,t) \quad (8)$$

由此选出最短的正确响应。随后，使用所得数据集 $\{(u,\ t,\ o^*)\}$ 进行监督微调，以训练 π_θ 朝着简洁推理的方向发展。

长度约束强化学习（LCRL）。遵循 SAFE-RS，我们通过复合奖励优化效率 $R_{LC} = \alpha R_{acc} + \beta R_{fmt} + \lambda R_{eff}$，其中纳入准确性（式3）和格式（式4）目标。我们设定 $\lambda = 1$。效率奖励 R_{eff} 会惩罚令牌超量$E = \max(0,\ |\tau| - L_{threshold})$：

$$R_{eff} = -\min\left(\max\left(\frac{\ln(E+10)}{\ln(B)}, 0.1\right), 1\right)\cdot P^{max} \quad (9)$$

其中，$B = 1000$ 控制曲率，P^{max}设定幅度。这种对数缩放能够抑制冗长表达，同时不妨碍必要的推理，从而确保快速、可靠地进行欺诈检测。

3.4 实时微调

阶段 3 可对序列音频 $u_{1:i}$进行动态风险评估。在每个轮次 i，模型都会识别对话阶段（早期、后期、最终阶段）。通过提示工程，我们引导模型：**(1)** 允许在早期作出判断，**(2)** 在后期阶段形成结论，以及 **(3)** 要求在最终阶段作出决定。我们通过以下方式训练阶段感知：

$$R_{phase} = r_b \cdot \mathbb{I}_{corr} - r_p \cdot \mathbb{I}_{fail} \quad (10)$$

其中，$\mathbb{I}_{corr}$ 表示正确的阶段识别，$\mathbb{I}_{fail}$ 表示在最终阶段犹豫不决，r_b，r_p 分别表示相应的奖励和惩罚幅度。总奖励为 $R_{total} = \alpha R_{acc} + \beta R_{fmt} + \eta R_{depth} + \delta R_{phase}$。我们设置 $\delta = 5$ 和 $\eta = \mathbb{I}_{\text{non-SFT}}$。这一配置兼顾了准确性（$R_{acc}$）与符合阶段特点的决策时机（$R_{phase}$）。

Type	模型	分类				质量评估				Fin.
		Sce.	Fra.	FT.	AVG	Log.	Pra.	Exp.	SUM	
自动语音识别+大语言模型	GLM4-9B-Chat	75.10	46.91	82.22	68.08	1.61	1.43	2.20	5.24	51.14
	InternLM2.5-20B	78.34	36.67	**85.42**	66.81	1.99	1.93	2.43	6.35	50.21
	Qwen2.5-72B	78.31	**51.44**	81.24	**70.33**	2.21	2.16	2.70	7.07	**52.87**
	豆包 1.5 Pro	71.14	36.11	82.25	63.17	1.94	1.75	2.60	6.29	47.48
	DeepSeek V3	**88.53**	14.62	66.71	56.62	**2.32**	**2.34**	**2.85**	**7.51**	42.59
自动语音识别+学习率M	Deepseek R1	83.60	79.25	85.16	82.67	**3.18**	**3.26**	**3.50**	**9.94**	62.17
LALM	GLM4-9B-Voice	0.00	26.83	38.33	21.72	0.89	0.64	0.65	2.18	16.33
	Gemini-2-Flash	80.51	59.61	83.53	74.55	**2.25**	**2.29**	**2.72**	**7.26**	56.03
	GPT-4o	80.25	50.00	**86.26**	72.17	2.12	2.10	2.56	6.78	54.24
	Step-Audio-Chat	76.35	40.65	79.71	65.57	1.64	1.62	2.01	5.27	49.27
	Qwen2-Audio-7B-Instruct	70.22	58.51	20.48	49.74	1.51	1.42	1.96	4.89	37.38
	AntiFraud-Qwen2Audio	**81.31**	**84.78**	82.91	**83.00**	2.06	2.07	2.31	6.44	**62.36**
大型音频语言模型	**SAFE-RL**	81.57	**90.20**	**87.25**	**86.34**	**2.5**	2.64	2.97	**8.11**	**64.89**
	SAFE-RS	81.60	**89.61**	86.39	85.87	2.45	2.6	**2.99**	8.04	64.53
	SAFE-LS	**84.64**	**89.61**	**88.23**	**87.49**	2.49	**2.65**	2.97	**8.11**	**65.76**

表 1：模型在 TeleAntiFraud-Bench 上的性能。**红色**数值表示最先进水平的结果，**蓝色**数值表示次优性能，**粗体**数值表示相应模型类型中的最佳性能。

4 实验

4.1 实验设置

数据集。我们使用**TeleAntiFraud-28k**（Ma 等人，2025年）（28,511 对样本）开展三项任务：七分类场景、二分类欺诈检测和七分类欺诈类型识别。对于强化学习，我们仅使用原始音频和上下文，不使用推理标注。在实时微调中，轮次 $u_{1:n}$ 被划分为早期（$i < n/2$）、后期（$n/2 \leq i < n$）和最终（$i = n$）阶段。评估在保持分布的**TeleAntiFraud-Bench**(Ma 等人，2025年）上进行。

基线方法。我们将本方法与专有模型（GPT-4o（Hurst 等人，2024年）、Gemini-2-Flash（DeepMind，2024年）、豆包-1.5（TEAM，2025年）以及开源基线进行比较，其中包括 Deepseek V3/R1（Liu 等人，2024年；Guo 等人，2025年）、GLM-4-Voice（Zeng 等人，2024年）、Step-Audio（Huang 等人，2025年）以及 Qwen2-Audio 变体（Chu 等人，2024年；Ma 等人，2025年）。这一选择涵盖了不同规模的推理专用模型、自动语音识别+大语言模型级联，以及端到端多模态架构。

评估指标。我们报告场景（Sce.）、欺诈（Fra.）和类型（FT.）任务的加权F1，并取平均值作为 **平均值**。推理质量按逻辑严谨性（Log.）、实用价值（Pra.）和表达（Exp.）评分（0-5），总分为**总分**（0-15）。最终指标为金融平均值 $= 0.75 \cdot$平均值$+0.25 \cdot$（总分$/15$）。**实现细节**。我们的骨干模型是 AntiFraud-Qwen2Audio（Ma 等人，2025年），这是 Qwen2-Audio-7B-Instruct（Chu 等人，2024年）。训练流程依次采用基于规则的强化学习（SAFE-RL）、拒绝采样（SAFE-RS）和长度约束强化学习（SAFE- LS）。详细的超参数见附录 C。

4.2 SAFE-QAQ 在电信欺诈分析中的有效性

跨任务的性能提升。表 1中的实验结果揭示了 TeleAntiFraud-Bench 各项任务中一致的性能层级：我们的 SAFE-QAQ 系列模型相较于通用基线和专用的 AntiFraud-Qwen2Audio 基础模型，均展现出逐步提升的性能。在场景分类（*Sce.*）中，尽管 Deepseek V3（88.53）等大规模大语言模型凭借文本理解能力，在这项很大程度上基于语言的任务中占据优势，但我们的 SAFE-LS（84.64）显著优于其前身 AntiFraud-Qwen2Audio（81.31）。二者采用相同架构，这一结果证实，我们的强化学习框架即使在基础模型已具备较强能力的任务上，也能提升性能。鉴于参数量为 8B 的 AntiFraud-Qwen2Audio 经过领域微调后已超越大多数自动语音识别+大语言模型基线，这一 3.33 分的提升尤其值得关注。

最显著的进展体现在欺诈检测（*Fra.*）方面，从基础模型到最终系统的演进轨迹清晰可见。通用型 Qwen2-Audio-7B-Instruct 仅达到 58.51，而其经监督微调

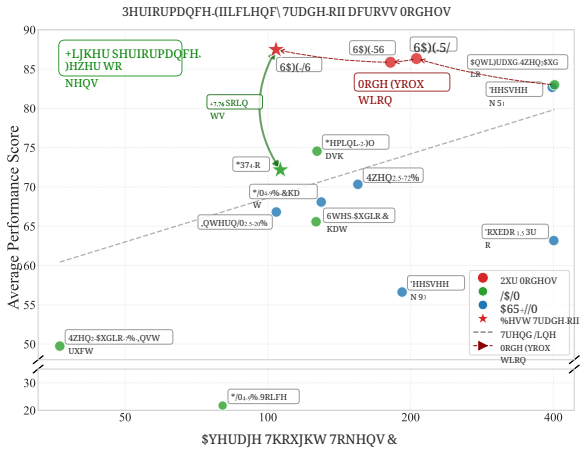

图3：性能-效率权衡：平均思考令牌数与平均分类性能的散点图。越靠近左上角的模型，在效率（思考令牌更少）和性能（分类得分更高）之间的平衡越好。基线方法和我们的模型中权衡最佳的点以星形标记突出显示。

增强的版本 AntiFraud-Qwen2Audio 通过慢思考适配达到 84.78，已经超越了 Deepseek R1（79.25）等专门的基于文本的模型。随后，我们的 SAFE-RL 通过基于规则的强化学习将这一成绩进一步提升至 90.20，实现了 5.42 分的绝对提升，展现出该方法识别细微多模态欺诈模式的卓越能力。从通用型大型音频语言模型，到领域适配的监督微调模型，再到经过强化学习优化的系统，这一演进过程验证了各训练阶段的互补价值；最终，SAFE-LS 在欺诈类型分类（$FT.$）中达到 88.23，甚至超越 GPT-4o（86.26），并为细粒度多模态分析确立了新的基准。

推理能力提升。 全面的指标进一步验证了这一方法。尽管 AntiFraud-Qwen2Audio（62.36 金融）通过多模态微调，表现已超过 Deepseek R1（62.17）等专用文本模型，但我们的 SAFE-LS（65.76）凭借完整的优化流程达到了新的最先进水平。最终提升的 3.40 分，反映出各项能力均衡进步；其质量评估分数（总分 8.11）已接近 Deepseek R1（总分 9.94）等专用推理模型，尽管 SAFE-LS 仅使用 8B 参数，而 Deepseek R1 使用了 671B 参数。这一效率提升得益于我们的强化学习框架，

模型	Sce.	Fra.	FT.	AVG	Dur.	轮次
base	70.22	58.51	20.48	49.74	*48.31秒*	*6.36*
采用监督微调	81.31	84.78	82.91	83.00	*48.31秒*	*6.36*
SAFE-RL	81.57	90.20	87.25	86.34	*48.31秒*	*6.36*
SAFE-RS	81.60	89.61	86.39	85.87	*48.31秒*	*6.36*
SAFE-LS	84.64	89.61	88.23	87.49	*48.31秒*	*6.36*
SAFE-Real	91.40	88.93	77.56	85.96	**8.98秒**	**1.25**

表 2：SAFE-Real 与基线方法对比

该框架系统性地促进慢思考过程，增强模型的逻辑推理、实践判断和表达质量。这些结果共同表明，传统方法虽在特定能力方面表现出色，但我们的端到端多模态框架在检测准确性、分类精确率和推理质量必须协同发挥作用的场景中，能够实现更优异的现实世界性能。

4.3 性能-效率权衡

在实际的欺诈检测系统中，计算效率会直接影响运营成本和响应时间——更短的推理链能够在实时通话期间更快地完成欺诈识别，同时降低基础设施开支。图3 展示了不同检测系统之间的性能-效率权衡，并通过我们提出的思考效率指标（TEM $=AVG/\log(|\tau|)$进行量化。该指标基于推理令牌数量，从与硬件无关的角度衡量效率；附录A中提供了系统级性能的详细分析，包括实际运行延迟和吞吐量。散点图根据每个模型的平均推理令牌长度（$|\tau|$，以对数尺度表示）与平均分类性能（AVG）对模型进行定位，揭示了根本性的架构差异和优化轨迹。位于左上区域附近的系统，在计算效率（更短的推理链）与检测准确率（更高的 F1 分数）之间实现了更优的平衡。

音频的优势。 大型音频语言模型架构（绿色圆圈）系统性地优于自动语音识别+大语言模型基线方法（蓝色 圆圈），其思考效率指标得分平均分别为 33.20 和 29.72。这种效率优势源于大型音频语言模型原生的多模态处理能力；这种能力不仅消除了级联式自动语音识别+大语言模型流水线中固有的误差累积，还能捕捉超越语音内容本身的音频特定语义信息，包括丰富的副语言信号。**欺诈推理中的缩放定律**。 我们观-

模型	Sce.	Fra.	FT.	AVG
SAFE-RL	81.57	90.20	87.25	**86.34**
无监督微调	73.37	85.70	82.91	80.66
无监督微调，使用 R_{depth}	77.25	86.97	81.94	82.05
SAFE-RS	81.60	89.61	86.39	**85.87**
无监督微调	79.81	87.14	83.82	83.59
SAFE-LS	84.64	89.61	88.23	**87.49**
无监督微调	82.89	91.42	87.07	87.13
无 RS	82.39	90.43	86.10	86.31

表3：消融实验：我们模型的性能

模型	Log.	Pra.	Exp.	Sum	令牌
SAFE-RL	2.50	2.64	2.97	8.11	205.76
不使用监督微调	1.93	1.92	2.39	6.24	31.46
不使用监督微调 w R_{depth}	2.10	2.07	2.59	6.76	212.72
SAFE-RS	2.45	2.60	2.99	8.04	181.33
无监督微调	2.09	2.09	2.60	6.78	173.78
SAFE-LS	2.49	2.65	2.97	8.11	104.02
无监督微调	2.17	2.25	2.64	7.06	74.05
无推理采样	2.23	2.31	2.78	7.32	106.20

表4：消融实验：我们模型的推理质量和令牌效率

察到推理复杂度与性能增益之间存在类缩放定律关系：增加推理令牌数量会带来检测准确率的对数改进。我们的分析揭示出极为一致的缩放模式：比较 GPT-4o 与 Gemini-2-Flash（$\Delta log(|\tau|) = 0.0773$）以及 GLM4-9B-Chat 与 Qwen2.5-72B（$\Delta log(|\tau|)$ =0。0772）时，我们发现几乎相同的性能增益（$\Delta AVG = 2.38$，而 2。25 分别），其比率 $\frac{\Delta log(|\tau|)}{\Delta AVG}$ 在不同模型族之间保持稳定（0.0325±0.0018）。这种缩放行为表明，欺诈推理任务呈现出与大语言模型预训练中观察到的现象相似的基本动态，尽管我们的强化学习框架最终通过定向优化打破了这一模式。

多阶段优化的有效性。我们的 SAFE优化路径（深红色轨迹）在保持性能优势的同时，展现出系统性的效率提升。我们的优化从慢思考型 AntiFraud-Qwen2Audio （ TEM=31.87 ）开始，SAFE-RL 中基于规则的强化学习将平均推理令牌数减少了 48.87%，同时使 $F1_{avg}$ 提升了 3.34 分（TEM=37.32）。随后，拒绝采样微调（SAFE-RS）又将令牌数减少了 11.87%，之后通过长度约束强化学习，在 SAFE-LS 中完成最终优化（TEM=43.38）。这三阶段的迭代优化使 TEM 比原始基础模型高出 36.12%，最终通过协同压缩推理过程并提升性能，使 TEM 超越 GPT-4o 7.76 分。

高性价比的SOTA性能。 最右侧的簇包含 Deepseek R1（TEM=16.03）等专用推理模型，它们通过大量生成令牌（平均 >397 个令牌）来实现有竞争力的准确性。尽管这些系统的性能已接近上限，但其运营成本阻碍了它们在现实世界中的部署。我们的 SAFE-LS 仅使用同类模型所需令牌的 25.85%，便实现了最先进的准确率（88.23 AVG）。这种高效性使我们已部署的系统每天能够处理超过 70,000 通筛查通话，通过自动干预显著减轻人工审核负担并防止财务损失。这为欺诈检测树立了新的实用基准。

4.4 实时检测性能

表2表明，SAFE-Real 平均仅需 8.98 秒即可实现实时检测（比非实时模型快 81.4%），同时保持稳健的欺诈检测性能（88.93 F1 分数）。尽管与 SAFE-LS 相比，其欺诈类型分类准确率下降了 12.2%（77.56 F1 分数），但这一权衡在实际操作中是合理的：在现实世界的欺诈防范中，及早检测比精准分类更为重要，因为及时阻止正在进行的诈骗可以避免财务损失。该模型出色的场景分类表现（91.40 F1 分数）以及平均仅 1.25 轮的极低对话轮次，使其实时通话拦截极为有效。

4.5 消融研究

表 3和 4 中的消融研究展示了监督微调预训练的关键作用，以及我们的 SAFE框架带来的渐进式改进。移除监督微调预训练会导致所有指标显著下降（例如，SAFE-RL 的性能下降 5.68 分至 80.66，推理质量下降 1.87 分至 6.24），尽管在没有监督微调预训练的情况下加入深度推理奖励（R_{depth}）可以部分缓解这些损失（分别提升至 82.05 和 6.76）。我们的令牌高效变体（SAFE-RS/SAFE-LS）在使用显著

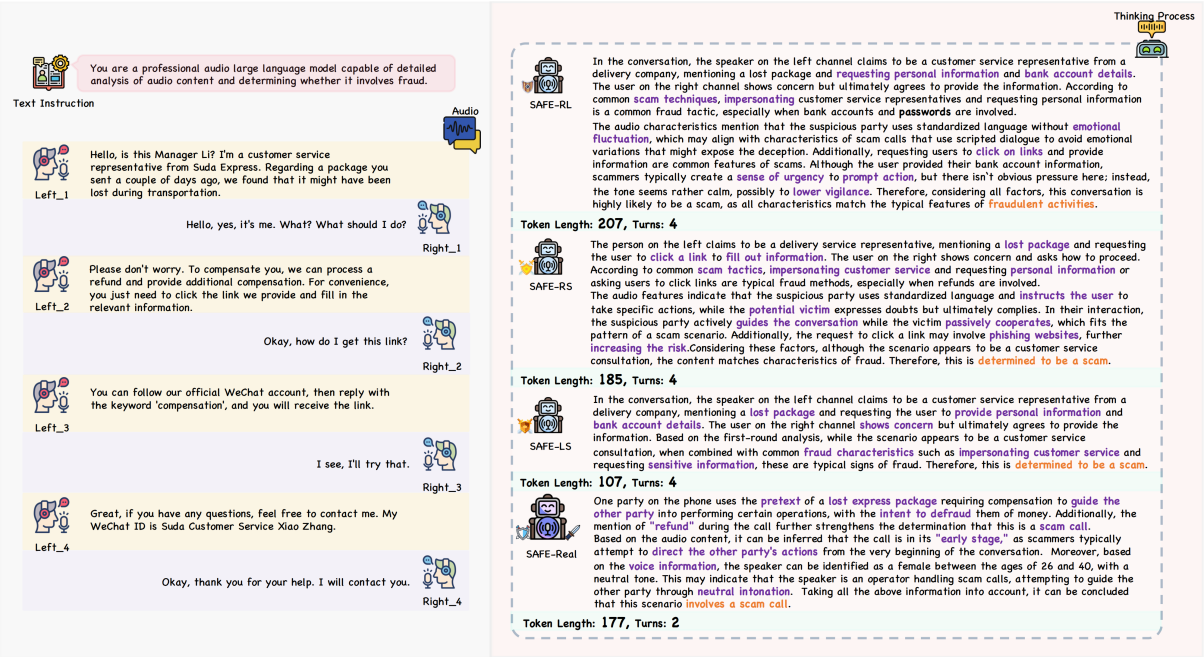

图 4：模型输出案例研究：文本指令和音频输入（为便于理解，左侧展示自动语音识别结果），以及 SAFE-QAQ 系列的推理过程（右侧）。关键推理要点以紫色突出显示，推理结果则以橙色标注。

更少令牌（181.33/104.02，相比之下为 205.76）的同时，仍保持较高的性能（85.87/87.49）和推理质量（分别为 8.04/8.11）。然而，它们仍然受益于监督微调预训练的基础能力：移除监督微调预训练后，SAFE-RS/SAFE-LS 的性能降至 83.59/87.13，推理质量也降低至 6.78/7.06。这证实了监督微调预训练在建立基础能力方面的关键作用，而后续强化学习阶段是在此基础上进行增强，而非取而代之。值得注意的是，当我们在 Long2Short 阶段移除拒绝采样（SAFE-LS w/o RS）时，性能会下降（从 87.49 降至 86.31），推理质量也随之降低，这表明拒绝采样微调对于在后续阶段提升效率的同时保持性能是必不可少的。

5 案例研究

图4 展示了我们的模型在典型的“包裹丢失退款”骗局场景中的推理过程。从 SAFE-RL 到 SAFE-RS 和 SAFE-LS，由于 Long2Short 优化，模型推理过程的长度逐渐缩短（从 207 个令牌缩短至 107 个令牌）。与此同时，关键推理要点的密度有所提高，表明 Long2Short 优化使推理更加高效，并提升了模型效率。为分析通话当前所处阶段，SAFE-Real 采用了长度适中的推理过程。值得注意的是，SAFE-Real 仅通过两轮对话音频，就实现了可解释的欺诈检测，凸显了其高效率。关键的是，我们的模型直接从原始音频中提取重要的副语言线索（口音、情绪、语音语调），而这些线索会在自动语音识别转写文本中丢失，从而揭示出平静语气与紧迫意图之间的微妙不一致。

6 结论

我们提出了 SAFE-QAQ，这是一种通过强化学习训练的端到端慢思考音频文本欺诈检测框架。通过结合 GRPO、基于规则的奖励、Long2Short 优化和实时学习，我们的模型在 TeleAntiFraud-Bench 上取得了最先进的性能（F1 分数为 88.23），同时效率显著提升（思维链缩短 48.87%，速度提升 81.4%）。除了学术指标之外，SAFE-QAQ 已成功部署于生产流水线，每天处理超过 70,000 通电话，并有效减轻人工审核负担、避免财务损失。本研究验证了多模态慢思考架构既能保持稳健性，又能实现实用的高效率，为应对现实世界的安全挑战提供了可扩展的解决方案。

局限性

尽管 SAFE-QAQ 展现出卓越的性能，但我们的实验范围不可避免地受到限制，因为 TeleAntiFraud-28k 目前是唯一适用于训练大型音频语言模型进行欺诈检测的开源数据集。数据稀缺使我们难以在高度多样化的欺诈场景或意外声学条件下，更全面地评估模型的泛化能力。因此，尽管我们的模型表现出较强的抗噪性，仍需进一步验证，以确保其在存在严重干扰或信号衰减的极端真实环境中的稳健性。

伦理声明

本研究严格遵循数据隐私标准。在实验验证中，我们使用了匿名化的 TeleAntiFraud-28k 数据集，确保不会泄露任何个人可识别信息（PII）。关于现实部署，我们与电信运营商合作，通过保护隐私的中间号码服务来实施系统。这种方法会屏蔽实际电话号码，所有数据处理均由企业授权，且仅用于反欺诈质量检查。我们强调，SAFE-QAQ 专为防御用途而设计，并要求持续监控，以防止滥用并减轻实际应用中可能出现的算法偏见。

参考文献

Josh Achiam、Steven Adler、Sandhini Agarwal、Lama Ahmad、Ilge Akkaya、Florencia Leoni Aleman、Diogo Almeida、Janko Altenschmidt、Sam Altman、Shyamal Anadkat 等另外 1 位作者。2023年。GPT-4 技术报告。*arXiv* 预印本 *arXiv:2303.08774*。

Joymallya Chakraborty、Wei Xia、Anirban Majumder、Dan Ma、Walid Chaabene 和 Naveed Janvekar。2024年。Detoxbench：面向多任务欺诈与滥用检测的大语言模型基准测试。*arXiv* 预印本 *arXiv:2409.06072*。

Chen-Wei Chang、Shailik Sarkar、Shutonu Mitra、Qi Zhang、Hossein Salemi、Hemant Purohit、Fengxiu Zhang、Michin Hong、Jin-Hee Cho 和 Chang-Tien Lu。2024年。揭示大语言模型漏洞：对抗性诈骗检测与性能。载于*2024*年 *IEEE* 大数据国际会议（*BigData*），第 3568–3571 页。IEEE。

Xingyu Chen、Jiahao Xu、Tian Liang、Zhiwei He、JianhuiPang、DianYu、Linfeng Song、QiuzhiLiu、Mengfei Zhou、Zhuosheng Zhang 及另外 1 位作者。2024 年。不要想得太多： 2+ 3=? 关于 o1类大语言模型的过度思考。*arXiv* 预印本 *arXiv:2412.21187*。

Yunfei Chu、Jin Xu、Qian Yang、Haojie Wei、Xipin Wei、Zhifang Guo、Yichong Leng、Yuanjun Lv、Jinzheng He、Junyang Lin 及另外 1 位作者。2024 年。Qwen2-Audio 技术报告。*arXiv* 预印本 *arXiv:2407.10759*。Google DeepMind。2024 年。Introducing Gemini 2.0：面向智能体时代的新 AI 模型。

Abhimanyu Dubey、Abhinav Jauhri、Abhinav Pandey、Abhishek Kadian、Ahmad Al-Dahle、Aiesha Letman、Akhil Mathur、Alan Schelten、Amy Yang、Angela Fan 等。2024年。Llama 3 模型群。*arXiv* 预印本 *arXiv:2407.21783*。

Daya Guo、Dejian Yang、Haowei Zhang、Junxiao Song、Ruoyu Zhang、Runxin Xu、Qihao Zhu、Shirong Ma、PeiyiWang、Xiao Bi，以及其他1位作者。2025年。Deepseek-r1：通过强化学习激励大语言模型的推理能力。*arXiv* 预印本 *arXiv:2501.12948*。

Sihao Hu、Tiansheng Huang、Ka-Ho Chow、Wenqi Wei、Yanzhao Wu 和 Ling Liu。2024年。Zipzap：在区块链上高效训练用于大规模欺诈检测的语言模型。载于*2024*年 *ACM Web* 会议论文集，第2807–2816页。

Ailin Huang、Boyong Wu、Bruce Wang、Chao Yan、Chen Hu、Chengli Feng、Fei Tian、Feiyu Shen、Jingbei Li、Mingrui Chen，以及其他1位作者。2025年。Step-Audio：智能语音交互中的统一理解与生成。*arXiv* 预印本 *arXiv:2502.11946*。

Aaron Hurst、Adam Lerer、Adam P Goucher、Adam Perelman、Aditya Ramesh、AidanClark、AJOstrow、Akila Welihinda、Alan Hayes、Alec Radford 等人。2024年。GPT-4o 系统卡。*arXiv* 预印本 *arXiv:2410.21276*。

Aaron Jaech、Adam Kalai、Adam Lerer、Adam Richardson、Ahmed El-Kishky、Aiden Low、Alec Helyar、Aleksander Madry、Alex Beutel、Alex Carney 等人。2024年。OpenAI o1 系统卡。*arXiv* 预印本 *arXiv:2412.16720*。

Mingyu Jin、Qinkai Yu、Dong Shu、Haiyan Zhao、Wenyue Hua、Yanda Meng、Yongfeng Zhang 和 Mengnan Du。2024年。推理步骤长度对大语言模型的影响。载于*2024*年计算语言学协会 *ACL* 论文集，第 1830–1842 页。Daniel Kahneman。2011年。思考，快与慢。麦克米伦。Sukanth Korkanti。2024年。利用大语言模型和高级数据分析提升金融欺诈检测。载于*2024*年第二届自维持人工智能系统国际会议（*ICSSAS*），第 1328–1334 页。IEEE。

Aixin Liu、Bei Feng、Bing Xue、Bingxuan Wang、Bochao Wu、Chengda Lu、Chenggang Zhao、Chengqi Deng、Chenyu Zhang、Chong Ruan 等 1 人。2024年。Deepseek-v3 技术报告。*arXiv* 预印本 *arXiv:2412.19437*。

Pan Lu、Liang Qiu、Kai-Wei Chang、Ying Nian Wu、Song-Chun Zhu、Tanmay Rajpurohit、Peter Clark 和 Ashwin Kalyan。通过策略梯度进行动态提示学习，以实现半结构化数学推理。载于第十一届国际学习表征会议。

Zhiming Ma、Peidong Wang、Minhua Huang、Jin-peng Wang、Kai Wu、Xiangzhao Lv、Yachun Pang、Yin Yang、Wenjie Tang 和 Yuchen Kang。2025年。TeleAntiFraud-28k：用于电信欺诈检测的音频-文本慢思考数据集。载于第*33*届 *ACM* 多媒体国际会议论文集（*MM '25*），第 1–9 页，纽约，美国纽约州。ACM。

OpenAI。2025年。OpenAI o3-mini。

Biqing Qi、Xinquan Chen、Junqi Gao、Dong Li、Jianxing Liu、Ligang Wu 和 Bowen Zhou。2024年。交互式持续学习：快思考与慢思考。载于*IEEE/CVF*计算机视觉与模式识别会议论文集，第 12882–12892 页。

John Schulman、Filip Wolski、Prafulla Dhariwal、Alec Radford 和 Oleg Klimov。2017年。近端策略优化算法。*arXiv* 预印本 *arXiv:1707.06347*。

MaohaoShen、Guangtao Zeng、Zhenting Qi、Zhang-Wei Hong、Zhenfang Chen、Wei Lu、Gregory Wornell、Subhro Das、David Cox 和 Chuang Gan。2025a。Satori：通过自回归搜索，利用思维链行动强化学习增强大语言模型推理。*arXiv* 预印本 *arXiv:2502.02508*。

Zitong Shen、Sineng Yan、Youqian Zhang、Xiapu Luo、Grace Ngai 和 Eugene Yujun Fu。2025b。"它恰好在正确的时刻提醒了我"：探索基于大语言模型的电话诈骗实时检测。*arXiv* 预印本 *arXiv:2502.03964*。

Gurjot Singh、Prabhjot Singh 和 Maninder Singh。2025年。使用检索增强生成大语言模型进行高级实时欺诈检测。*arXiv* 预印本 *arXiv:2501.15290*。

´Maciej Swiechowski、Konrad Godlewski、Bartosz Saw- icki 和 Jacek Ma´ndziuk。2023年。蒙特卡洛树搜索：近期改进与应用综述。《人工智能评论》，56(3):2497– 2562。

豆包团队。2025年。豆包-1.5-pro。

Omkar Thawakar、Dinura Dissanayake、Ketan More、Ritesh Thawkar、Ahmed Heakl、Noor Ahsan、Yuhao Li、Mohammed Zumri、Jean Lahoud、Rao Muhammad Anwer 等。2025年。Llamav-o1：重新思考大语言模型中的逐步视觉推理。*arXiv* 预印本 *arXiv:2501.06186*。

Luong Trung、Xinbo Zhang、Zhanming Jie、Peng Sun、Xiaoran Jin 和 Hang Li。2024年。ReFT：通过强化微调进行推理。载于计算语言学协会第 *62* 届年会论文集（第 *1* 卷：长篇论文），第 7601–7614 页。

Guowei Xu、Peng Jin、Li Hao、Yibing Song、Lichao Sun 和 Li Yuan。2024年。LLaVA-o1：让视觉语言模型进行逐步推理。*arXiv* 预印本 *arXiv:2411.10440*。

Shu Yang、Shenzhe Zhu、Zeyu Wu、Keyu Wang、Junchi Yao、Junchao Wu、Lijie Hu、Mengdi Li、Derek F Wong 和 Di Wang。2025年。Fraud-R1：用于评估大语言模型抵御增强型欺诈和网络钓鱼诱导的稳健性的多轮基准。*arXiv* 预印本 *arXiv:2502.12904*。

Aohan Zeng、Zhengxiao Du、Mingdao Liu、Kedong Wang、Shengmin Jiang、Lei Zhao、Yuxiao Dong 和 Jie Tang。2024年。GLM-4-Voice：迈向智能且更具人类特质的端到端语音聊天机器人。*arXiv* 预印本 *arXiv:2412.02612*。

Xinyu Zhu、JunjieWang、Lin Zhang、Yuxiang Zhang、Ruyi Gan、Jiaxing Zhang 和 Yujiu Yang。2022。通过合作推理诱导语言模型解决数学应用题。*arXiv* 预印本 *arXiv:2210.16257*。

计算效率分析

为验证我们在理论上减少推理 token 数量的效果能否转化为实际的耗时节省，我们对推理延迟和吞吐量进行了详细的性能分析。

实验设置。 所有效率评估均在 NVIDIA A100 GPU 上进行。我们测量了三种迭代模型的性能：SAFE-RL、SAFE-RS 和 SAFE-LS。为确保测量结果一致且可复现，我们采用了贪心解码（温度 $= 0$）。我们报告中位延迟（$p50$）、第95百分位延迟（$p95$）以及系统整体吞吐量（每秒处理的样本数）。

结果。 如表 5所示，各优化阶段呈现出清晰的效率提升趋势。SAFE-RL 生成的推理链最长，因此延迟最高。拒绝采样阶段（SAFE-RS）通过过滤掉冗长的响应，带来了适度的改善。最值得注意的是，最终的长度约束强化学习阶段（SAFE-LS）实现了 $p50$ 916.2 毫秒的延迟和每秒 1.10 个样本的吞吐量。

与 SAFE-RL 基线相比，这相当于中位延迟降低了 $\approx 26.34\%$。这些结果证实，第 4.3 节讨论的“思考效率指标”（TEM）与现实部署指标高度相关。延迟大幅下降（从 1895.7 毫秒降至 1207.8 毫秒）也表明，SAFE-LS 对“无限循环”或过度思考的抵御能力显著更强，稳定性和鲁棒性更佳，因此更适用于对时效性要求较高的欺诈检测场景。

B 提示词

本研究使用的提示词旨在引导模型完成推理和决策的各个阶段。以下详细说明了它们的作用：

- 图 5 展示了用于训练三个任务第一阶段和第二阶段的提示词：场景分类、欺诈检测和欺诈类型分类。这些提示词经过结构化设计，以促进慢思考推理，使模型能够捕捉音频细节中的细微差异，例如声调波动、情绪压力和环境线索。

表 5：NVIDIA A100 上的推理延迟和吞吐量分析。对三个训练阶段的比较表明，我们的 Long2Short 优化（SAFE-LS）显著降低了延迟并提高了吞吐量。

模型	p50	p95	吞吐量
SAFE-RL	1243.9	1895.7	0.76
SAFE-RS	1204.0	1263.4	0.85
SAFE-LS	**916.2**	**1207.8**	**1.10**

- 图6、图 7和图8 展示了实时检测阶段所采用的提示词。这些提示词会根据对话阶段（早期、中期或最终阶段）动态调整，在考虑可用信息是否充足的同时，确保及时且准确地进行欺诈检测。

- 图9 展示了为评估模型推理过程质量而设计的提示词。这些提示词侧重于评估逻辑严谨性、实用价值和表达质量，为模型性能提供了全面的评估框架。

C 超参数设置与敏感性分析

为确保可复现性并促进后续研究，我们全面而详细地描述了超参数配置，包括具体选择背后的理由，以及在开发 SAFE-QAQ 过程中开展的敏感性分析。

C.1 实现平台

所有实验均在配备 4 张 NVIDIA A100（80GB）GPU 的高性能计算集群上进行。我们采用全局批次大小 12（通过每设备批次大小 $bs = 3$ 和梯度累积实现）。训练框架基于 ms-swift[1]，构建，并针对大型音频语言模型微调进行了优化。

C.2 配置依据

我们的超参数选择策略兼顾训练稳定性、推理效率和任务性能。具体配置分类如下：

[1] https://github.com/modelscope/ms-swift

奖励权重配置。 组合奖励函数包含多个组成部分（$\alpha, \beta, \delta, \lambda$）。我们根据目标的优先级确定其取值:

- **准确性优先（$\alpha = 5,\ \beta = 1$）：** 我们将比例设为 $\alpha:\ \beta = 5:1$。对 α 赋予较大的权重，可确保模型优先考虑最终分类的正确性（准确性奖励），而非仅仅满足结构合规性（格式奖励）；同时，$\beta = 1$ 仍足以引导解析器。

- **阶段感知（$\delta = 5$）：** 对于实时微调，准确的阶段识别对于及时干预至关重要。我们将 $\delta = 5$设为与准确性权重 α相同，以强调在实时场景中，识别正确的对话阶段与欺诈检测本身同等重要。

- **效率平衡（$\lambda = 1$）：** 我们将 $\lambda = 1$设为推理长度引入正则化项。选择这一数值旨在控制冗长程度，同时避免其过度削弱准确性奖励，防止模型为了追求简洁性而牺牲必要的推理深度。

强化学习（GRPO）参数。我们采用组相对策略优化（GRPO）算法。

- **稳定性因素（$\epsilon = 0.2,\ \beta_{KL} = 0.04$）：** 我们遵循 ms-swift 框架推荐的默认设置。具体而言，裁剪系数 $\epsilon = 0.2$ 和 KL 散度系数 $\beta_{KL} = 0.04$ 对于防止策略崩溃至关重要。在初步实验中，我们尝试移除 KL 惩罚项（即 $\beta_{KL} = 0$），结果导致严重的训练不稳定性和模式崩溃。因此，我们保留了稳健的默认值。

- **组大小（$G = 9$）：** 我们将组大小设为 9。该值在计算开销与梯度方差降低之间取得了平衡，确保在 GPU 内存有限的情况下稳定收敛。

生成与长度约束。

- **采样策略：** 为了在探索过程中平衡生成多样性和质量，我们采用核采样，使用 $top_p = 0.9$, $top_k = 50$，温度设为 0。9。对于拒绝采样，我们将候选项数量设为 $K = 16$，以确保充分覆盖解空间。

- **长度阈值：** 最大长度阈值 $L_{max} = 200$ 以及针对 R_{depth} 和 R_{eff} 的阈值$L_{threshold} = 200$，是根据 TeleAntiFraud-28k 数据集中推理链的统计分布确定的。 $P^{max} = 5$ 用于限制惩罚幅度，防止梯度过大而导致策略不稳定。

C.3 超参数敏感性分析

为验证我们对学习率的选择——它是影响强化学习收敛的关键因素——我们使用 SAFE-RL（不含 SFT）模型，在 TeleAntiFraud-Bench 的一个合成子集上进行了网格搜索。结果汇总于表6。

表6：学习率（LR）对模型性能的敏感性分析。所选设置（$3e^{-5}$）在所有指标上实现了最佳平衡。

学习率	Sce.	Fra.	FT.	AVG
$1e^{-5}$	85.34	72.86	75.36	77.85
$\mathbf{3e^{-5}}$ **（我们的方法）**	84.31	**88.70**	75.74	**82.92**
$5e^{-5}$	85.12	78.10	**76.56**	79.93

如观察结果所示，学习率为 $3e^{-5}$ 时，平均 F1 分数最高（平均值：82.92）。较低的学习率（$1e^{-5}$）导致欠拟合，尤其是在欺诈检测（分数）任务中；而较高的学习率（$5e^{-5}$）则使性能下降，这可能是由于优化超调所致。因此，所有主要实验均采用 lr =$3e^{-5}$。

Prompt for Non-Real-Time

Task_1 Description:
Please analyze in detail which of the following types the call scenario belongs to based on the audio content you heard:
- Food ordering service
- Customer service inquiry
- Appointment service
- Transportation inquiry
- Daily shopping
- Ride-hailing service
- Food delivery service

Input:

Call record

Output Format:

Please provide your judgment result in the following format:
<think>
...(Detailed analysis process)
</think>
<answer>
```
{
"scene": "<scene_type>",
"reason": "<reason_for_judgment>",
"confidence": <confidence_level>
}
```
</answer>
- `scene`: A string indicating the identified call scenario type, must be chosen from the following closed set: `["Food ordering service", "Customer service inquiry", "Appointment service", "Transportation inquiry", "Daily shopping", "Ride-hailing service", "Food delivery service"]`.
- `reason`: A string providing a brief explanation of the reason for your judgment.
- `confidence`: A float indicating your confidence level in the judgment, ranging from 0 to 1, with 1 indicating complete confidence.

Task_2 Description:
You are a professional audio large language model capable of analyzing audio content in detail and determining whether it involves fraud. Based on the following information, provide your judgment:
1. The call scenario analyzed in the first round.
2. The audio content.

Output Format:

Please output your judgment result in the following format:
<think>
...(Detailed analysis process)
</think>
<answer>
```
{
  "reason": "<reason_for_judgment>",
  "confidence": <confidence_level>,
  "is_fraud": <true/false>
}
```
</answer>
- `reason`: A string briefly explaining the reason for your judgment.
- `confidence`: A float indicating your confidence level in the judgment, ranging from 0 to 1, with 1 indicating complete confidence.
- `is_fraud`: A boolean value indicating whether the audio involves fraud. `true` indicates fraud, and `false` indicates no fraud.

Task_3 Description:
You are a professional audio large language model capable of analyzing audio content in detail and determining the type of fraud involved. Based on the following information, provide your judgment:
1. The call scenario analyzed in the first round.
2. The second-round analysis of whether the call involves fraud.
3. The audio content.

Output Format:

Please provide your judgment result in the following format:
<think>
...(Detailed analysis process)
</think>
<answer>
```
{
  "fraud_type": "<fraud_type>",
  "reason": "<reason_for_judgment>",
  "confidence": <confidence_level>
}
```
</answer>
- `fraud_type`: A string indicating the identified fraud type, which must be chosen from the following closed set: `["Investment Fraud", "Phishing Fraud", "Identity Theft", "Lottery Fraud", "Bank Fraud", "Kidnapping Fraud", "Customer Service Fraud", "Email Fraud"]`.
- `reason`: A string briefly explaining the reason for your judgment.
- `confidence`: A float indicating your confidence level in the judgment, ranging from 0 to 1, with 1 indicating complete confidence.

图5：非实时提示。

Prompt for Real-Time Scenario Classification

````
**Task Description:**
Please analyze the content of the audio you heard in detail and determine which of the following types the call scenario belongs to, as well as the stage of the call.

**Scene Types:**

- Food Ordering Service
- Customer Service Inquiry
- Appointment Service
- Transportation Inquiry
- Daily Shopping
- Ride-hailing Service
- Food Delivery Service

**Input:**

Audio clip of the call

**Output Format:**

Please strictly follow the format below to output your judgment result:

```
<think>
...(Detailed analysis process, including whether there is enough information to support the conclusion)
</think>
<answer>
...(Specific judgment content, format shown below)
</answer>
```

1. **`<think>` Section:**
   - Detailed analysis of the audio content, including:
     - Understanding of the call scenario
     - Extraction of key information
     - Evaluation of whether there is sufficient information to support the conclusion
     - Assessment of the current stage of the call
   - Must be wrapped with `<think>` and `</think>` tags.

2. **`<answer>` Section:**
   - Output the specific judgment result based on the analysis.
   - Must be wrapped with `<answer>` and `</answer>` tags.
   - Output a JSON object in the following format:
     ```json
     {
       "conversation_stage": "<stage>",
       "scene": "<scene_type>|null",
       "reason": "<reason_for_judgment>",
       "confidence": <confidence_level>
     }
     ```
     - `conversation_stage`: A string indicating the stage of the call, must be chosen from the following closed set:
       - `"early_stage"` (The first half of the call, where the intent of the conversation has not been fully revealed)
       - `"late_stage"` (The latter half of the call, nearing the end of the conversation)
       - `"complete"` (Full call record)
     - `scene`: A string or null, indicating the identified call scenario type, must be chosen from the following closed set or be null: `["Food Ordering Service", "Customer Service Inquiry", "Appointment Service", "Transportation Inquiry", "Daily Shopping", "Ride-hailing Service", "Food Delivery Service", null]`.
       - When `conversation_stage` is `"complete"`, this field cannot be null.
       - When `conversation_stage` is `"late_stage"`, a judgment should be provided if possible.
       - When `conversation_stage` is `"early_stage"`, this field can be null.
     - `reason`: A string briefly explaining the reason for your judgment, including an analysis of the call stage.
     - `confidence`: A float indicating your confidence level in the judgment, ranging from 0 to 1, with 1 indicating complete confidence.

3. **Notes:**
   - The overall output format must be strictly followed:
     ```
     <think>
     ...(Detailed analysis process)
     </think>
     <answer>
     ...(Specific judgment content)
     </answer>
     ```
   - Both the `<think>` and `<answer>` sections must exist, and their order cannot be reversed.
   - The JSON object must contain all specified fields.
   - The judgment of the call stage should be explained in detail in the `<think>` section.
````

图6：实时场景分类提示词。

Prompt for Real-Time Fraud Detection

````
**Task Description:**
Based on the audio content you hear, analyze it in detail and determine whether it involves fraud and the stage of the call. Please provide your judgment based on the following information:

1. The call scenario analyzed in the first round.
2. The audio content.

**Output Format:**

Please strictly follow the format below to output your judgment results:

```
<think>
...(Detailed analysis process, including whether there is sufficient information to make a judgment)
</think>
<answer>
...(Specific judgment content, see format below)
</answer>
```

1. **`<think>` Section:**
   - Detailed analysis of the audio content, including:
     - Understanding of the call scenario
     - Extraction of key information
     - Determination of whether there is enough information to support the conclusion
     - Evaluation of the current stage of the call
   - Must be wrapped with `<think>` and `</think>` tags.

2. **`<answer>` Section:**
   - Based on the analysis results, output the specific judgment content.
   - Must be wrapped with `<answer>` and `</answer>` tags.
   - Output a JSON object in the following format:
     ```json
     {
       "conversation_stage": "<stage>",
       "reason": "<reason_for_judgment>",
       "confidence": <confidence_level>,
       "is_fraud": <true/false>|null
     }
     ```
     - `conversation_stage`: A string indicating the stage of the call, which must be chosen from the following closed set:
       - `"early_stage"` (The first half of the call, where the intent of the conversation has not been fully revealed)
       - `"late_stage"` (The second half of the call, nearing its conclusion)
       - `"complete"` (A full call record)
     - `reason`: A string briefly explaining the reason for your judgment.
     - `confidence`: A float indicating your confidence level in the judgment, ranging from 0 to 1, with 1 indicating complete confidence.
     - `is_fraud`: A boolean value or null, indicating whether the audio involves fraud. `true` indicates fraud, `false` indicates no fraud. If there is insufficient information, it should be null.
       - When `conversation_stage` is `"complete"`, this field cannot be null.
       - When `conversation_stage` is `"late_stage"`, a judgment should be provided if possible.
       - When `conversation_stage` is `"early_stage"`, it can be null.

3. **Notes:**
   - You must strictly adhere to the following overall output format:
     ```
     <think>
     ...(Detailed analysis process)
     </think>
     <answer>
     ...(Specific judgment content)
     </answer>
     ```
   - Both the `<think>` and `<answer>` sections must exist, and their order cannot be reversed.
   - The JSON object must contain all specified fields.
   - The judgment of the call stage should be explained in detail in the `<think>` section.
````

图7：实时欺诈检测提示词。

Prompt for Real-Time Fraud Type Classification

Task Description:

Based on the audio content you hear, analyze it in detail and determine the type of fraud involved as well as the stage of the call. Please provide your judgment based on the following information:

1. The call scenario analyzed in the first round.
2. The analysis of whether the call involves fraud in the second round.
3. The audio content.

Output Format:

Please strictly follow the format below to output your judgment:

```
<think>
...(Detailed analysis process, including whether there is enough information to support the conclusion)
</think>
<answer>
...(Specific judgment content, see format below)
</answer>
```

1. **`<think>` Section:**
   - Provide a detailed analysis of the audio content, including:
     - Understanding of the call scenario.
     - Extraction of key information.
     - Assessment of whether there is sufficient information to support the conclusion.
     - Evaluation of the current stage of the call.
   - Must be enclosed in `<think>` and `</think>` tags.

2. **`<answer>` Section:**
   - Based on the analysis results, output the specific judgment content.
   - Must be enclosed in `<answer>` and `</answer>` tags.
   - Output a JSON object in the following format:
     ```json
     {
       "conversation_stage": "<stage>",
       "fraud_type": "<fraud_type>|null",
       "reason": "<reason_for_judgment>",
       "confidence": <confidence_level>
     }
     ```
     - `conversation_stage`: A string indicating the stage of the call, must be chosen from the following closed set:
       - `"early_stage"` (The first half of the call, where the intent of the conversation has not been fully revealed).
       - `"late_stage"` (The latter half of the call, where the conversation is nearing its end).
       - `"complete"` (A full record of the call).
     - `fraud_type`: A string or null, indicating the identified fraud type. Must be chosen from the following closed set or null: `["Investment Fraud", "Phishing Fraud", "Identity Theft", "Lottery Fraud", "Bank Fraud", "Kidnapping Fraud", "Customer Service Fraud", "Email Fraud", null]`.
       - When `conversation_stage` is `"complete"`, this field cannot be null.
       - When `conversation_stage` is `"late_stage"`, a judgment should be provided if possible.
       - When `conversation_stage` is `"early_stage"`, this field can be null.
     - `reason`: A string briefly explaining the reason for your judgment, including an analysis of the call stage.
     - `confidence`: A float indicating your confidence level in the judgment, ranging from 0 to 1, with 1 indicating complete confidence.

3. **Notes:**
   - The overall output format must be strictly followed:
     ```
     <think>
     ...(Detailed analysis process)
     </think>
     <answer>
     ...(Specific judgment content)
     </answer>
     ```
   - Both the `<think>` and `<answer>` sections must exist, and their order cannot be reversed.
   - The JSON object must contain all specified fields.
   - The judgment of the call stage should be explained in detail in the `<think>` section.

图8：实时欺诈类型分类提示词。

Prompt for Evaluating the Thought Process

```
Please evaluate the model's reasoning process probabilistically based on three dimensions: **logical consistency, practicality, and clarity**. For each scoring criterion, calculate the probability of achieving the corresponding score, and compute the final score as the expected value.

---

#### **Input**
1. Model's reasoning process: {reasoning_process}
2. Model's final answer: {model_answer}
3. Reference answer: {reference_answer}
4. Reference reasoning process: {reference_reasoning}

---

### Scoring Rules
**1. Logical Rigor (5 points)**
  ☐ Complete reasoning chain without gaps (0-1 point)
    - Probability of earning 1 point: __%
  ☐ Reasonable and explicit key assumptions (0-1 point)
    - Probability of earning 1 point: __%
  ☐ Strictness of conclusion derivation (0-2 points)
    - Probability of earning 2 points: __%
    - Probability of earning 1 point: __%
  ☐ Advantage probability compared to reference reasoning (0-1 point)
    - Probability of earning 1 point: __%

**2. Practical Value (5 points)**
  ☐ Accuracy in identifying the essence of the problem (0-1 point)
    - Probability of earning 1 point: __%
  ☐ Effectiveness of the solution (0-2 points)
    - Probability of earning 2 points: __%
    - Probability of earning 1 point: __%
  ☐ Completeness of addressing requirements (0-1 point)
    - Probability of earning 1 point: __%
  ☐ Optimization probability compared to reference reasoning (0-1 point)
    - Probability of earning 1 point: __%

**3. Expression Quality (5 points)**
  ☐ Completeness of presenting key nodes (0-1 point)
    - Probability of earning 1 point: __%
  ☐ Clarity of expression (0-2 points)
    - Probability of earning 2 points: __%
    - Probability of earning 1 point: __%
  ☐ Brevity of information (0-1 point)
    - Probability of earning 1 point: __%
  ☐ Expression advantage compared to reference reasoning (0-1 point)
    - Probability of earning 1 point: __%

---

### Probability Constraints
1. If the conclusion is incorrect:
  - Logical rigor: Probability of conclusion alignment drops to zero; upper limit for other items is 40%.
  - Practical value: Probabilities for problem identification and solution effectiveness drop to zero.
2. Logical leaps: Each instance reduces probabilities by 15-30%.
3. Missing evidence: Total probability for the corresponding dimension decreases by 20%.

---

### Output Template
**Logical Rigor**
- Complete reasoning chain without gaps (1-point probability): <__%>
- Reasonable and explicit key assumptions (1-point probability): <__%>
- Strictness of conclusion derivation (2 points → __% | 1 point → __%)
- Advantage probability compared to reference reasoning (1-point probability): <__%>
Expected score: <Calculation formula>

**Practical Value**
- Accuracy in identifying the essence of the problem (1-point probability): <__%>
- Effectiveness of the solution (2 points → __% | 1 point → __%)
- Completeness of addressing requirements (1-point probability): <__%>
- Optimization probability compared to reference reasoning (1-point probability): <__%>
Expected score: <Calculation formula>

**Expression Quality**
- Completeness of presenting key nodes (1-point probability): <__%>
- Clarity of expression (2 points → __% | 1 point → __%)
- Brevity of information (1-point probability): <__%>
- Expression advantage compared to reference reasoning (1-point probability): <__%>
Expected score: <Calculation formula>
```

图9：评估思维过程的提示词。