

返回 新闻中心 > 新闻详情

ACL 2025 | TeleAI 一冠一亚！多项成果推动大模型创新与应用

2025-07-29 10:15 中国电信人工智能研究院 (TeleAI)

本周末，中国计算机学会（CCF）推荐的 A 类会议、自然语言处理（NLP）领域顶级学术会议之一“ACL 2025”（第 63 届国际计算语言学年会）将正式召开。根据此前公布的论文录用结果，本届 ACL 共收到超过 8,000 篇投稿，相较去年的 4,407 篇增加了近一倍，难度再升级。

在中国电信集团 CTO、首席科学家、中国电信人工智能研究院（TeleAI）院长李学龙教授的指导下，TeleAI 多项研究成果被收录，其中包括 **1 篇 Main** 和 **2 篇 Findings**。TeleAI 科研团队还在会议前期举办的“国际语义评测（SemEval）”竞赛中斩获**一项冠军、一项亚军**。



练效率、推理能力、性能表现等展开基础研究，还在**表格问答、情感交流、代码开发**等方面为产业带来应用价值。

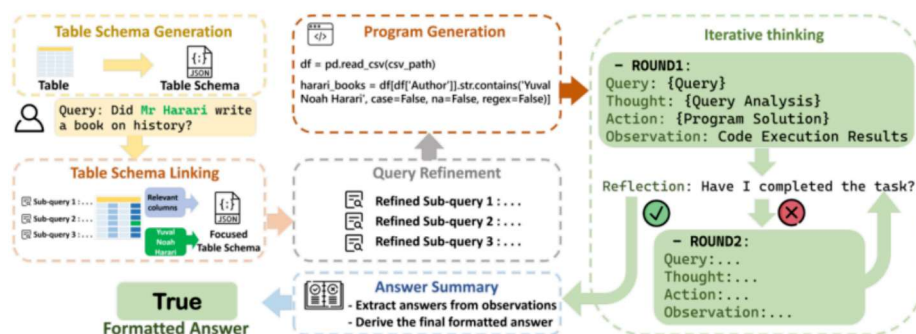
01

大规模表格问答

大规模表格问答是大语言模型的核心能力之一，在多种行业和领域有着重要的应用前景。它可以帮助投资者分析市场趋势，进行风险评估；辅助医生解读患者病例，处理临床数据；还能在客服场景提取客服话术，进行订单处理等。

本届 ACL 的 SemEval 挑战赛提出了基于表格数据回答自然语言问题的任务，包括表格规模大、语义不完整、实体歧义等挑战。此项任务目标推动表格数据理解的研究与发展，为自然语言处理在表格数据领域的应用提供支持。

TeleAI 团队提出了一个**基于大型语言模型且采用编程模式的系统性表格问答框架 TableReasoner**，创新引入了模式感知表格语义建模方法，从而大幅压缩冗余信息，使得能够处理超大规模的表格任务。



借助编程模块，TableReasoner 还可利用表格模式从整体角度理解表格内容以解决问题。框架设计了“意图理解-模式匹配-实体链接-Query 改写”的动作流，通过蒸馏与拒绝采样构建高质量推理数据，**实现大幅提升表格问答的准确率，改善幻觉问题。**

情绪与意图理解

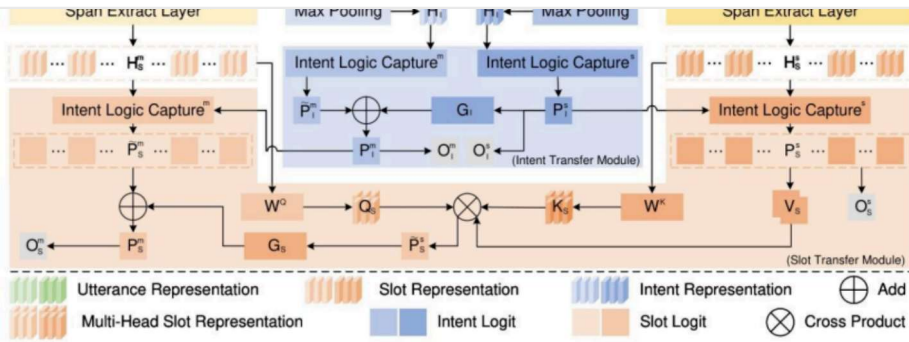
SemEval 挑战赛的另一项任务是聚焦基于文本的情感检测，旨在弥合当前文本情感检测领域存在的差距，推动该技术的快速发展。大语言模型的情感和意图理解在智能客服、情感陪伴等领域能够发挥巨大价值，具有重要的研究意义和应用前景。

TeleAI 团队针对多语言、多标签情感共存等复杂挑战，提出了**统一增强识别框架 EmotionLLM**，通过大语言模型驱动的伪标签生成与一致性筛选机制构建高质量增强样本，并结合上下文提示设计与多模型协同预测策略，实现情绪理解精度与鲁棒性的双重提升。

在多标签情感检测方面，EmotionLLM 基于上下文学习，能够动态优化提示（调整示例、加入情感线索），帮助模型捕捉情感细微差异。同时，通过大语言模型生成符合原始数据情感特征的合成数据，在过滤后保留标签一致样本，从而解决数据不平衡。

在情感强度预测方面，此框架针对单一情感（如愤怒）设计示例，并覆盖 0-3 级强度（无情感-低强度-中强度-高强度），以增强模型对强度差异的理解。此外，还将任务转为单情感多分类（每次预测一种情感的强度），以减少多情感干扰。

现有语言模型在处理语义相同但语言不同的输入时，性能存在差异。EmotionLLM 则**展现了出色的跨语言迁移能力和情感识别泛化能力**，在本次挑战赛的中、俄、德、葡四语种排名中名列全球第二。



03

网页开发与代码安全

随着生成式 AI 技术的快速发展，多模态大语言模型（MLLMs）展现出作为 AI 软件工程师的潜力，可执行复杂的网页开发任务。这类模型能够基于用户提供的 WebUI 图像自动生成前端代码项目（WebUI-to-Code），显著加速了网页开发的流程和效率。

然而，当前主流评测集通常只关注生成网页的视觉质量，缺乏对模型网页生成多维度子能力的系统评估。为了解决这一问题，TeleAI 团队受软件工程原理启发，提出了一个**新型评测数据集 WebUIBench**。

科研团队在该数据集构建过程中实施了严格的质量控制，专注于全面评估多模态大语言模型在网页开发场景中的四个核心能力维度，包括 WebUI 感知、HTML 编程、WebUI-HTML 理解、WebUI-to-Code 生成。

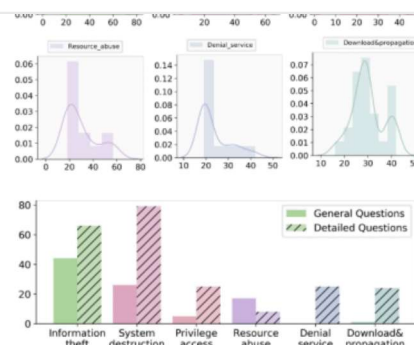
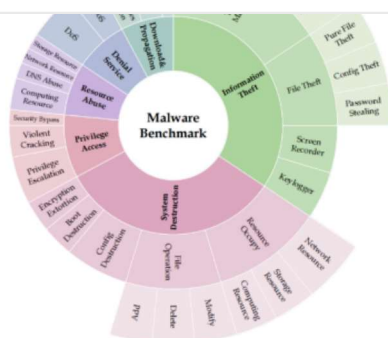


Figure 2: Task examples in the WebUI benchmark, from the **WebUI Perception**, **HTML Programming**, and **WebUI-HTML Understanding** and **WebUI-to-Code** task.

WebUIBench 为理解当前多模态大语言模型在网页开发领域的的能力边界提供了重要参考，也为未来研究提供了方向，包括如何增强模型的视觉定位能力和跨模态理解能力，以及如何平衡和提升不同维度的子能力，从而实现更高效的 AI 驱动的网页开发。

与此同时，在诸如自动化网页开发等代码生成应用快速发展的同时，通过大语言模型进行恶意代码生成，对网页、应用进行恶意攻击的案例屡见不鲜。这严重影响了生成式人工智能技术的发展环境和应用生态。

为此，TeleAI 团队还构建了一个 **MalwareBench 数据集**，包括 6 个大分类、29 个细分类、320 条初始恶意问题，能够有效模拟现实中针对大语言模型的恶意代码生成攻击情景，**为评估大语言模型的抵御能力提供标准化平台。**



此数据集不仅揭示了大语言模型在代码安全方面的挑战，同时也为后续研究提供了宝贵资源。研究人员可以在统一标准下深入探讨模型面向恶意软件生成这一具体方向时的安全性能，进而研发出更为有效的防御机制，促进相关技术的安全创新与应用。

除上述成果，在李学龙教授的带领下，TeleAI 实现了多模态大模型在代码生成领域的实际落地。继去年七月发布以来，星辰大模型-软件工厂持续迭代优化，先后支持多种前后端开发框架，并相继发布 WEB 端应用与 IDE 插件，在众多内外部用户中推广验证有效提升开发效率，收获大量好评。

相关论文:

LLMs Caught in the Crossfire: Malware Requests and Jailbreak Challenges (Main)

WebUIBench: A Comprehensive Benchmark for Evaluating Multimodal Large Language Models in WebUI-to-Code (Findings)

Towards Efficient LLM Grounding for Embodied Multi-Agent Collaboration (Findings)

TeleAI at SemEval-2025 Task 8: A Table Reasoning Framework with Language Models (Workshop)

Data Augmentation (Workshop)

上一篇 [WAIC 2025 | 入选 SAIL 奖 TOP30, TeleAI 国产大模型深度解析 \(第2弹\)](#)

下一篇 [解放双手，看见智慧！中国电信天翼AI眼镜震撼发布，你的第一视角智能伙伴来了！](#)