

麻紫嫣 Tzu-yen Ma

具有多模态大模型方向研究经历及多篇学术成果产出，学习与适应能力强

18612101325

maziyan@bupt.edu.cn

北京市



教育背景

北京邮电大学	计算机技术	硕士	2024.9-2027.7
第四届中国移动“梧桐杯”大数据创新大赛暨大数据创客马拉松全国总决赛一等奖，A类研究生创新创业项目总负责人，校级台港澳华侨研究生奖学金（名额1）			
北京邮电大学	数据科学与大数据技术	本科	2020.9-2024.7
国家励志奖学金、连续三年校级台港澳华侨本科奖学金			

科研/项目经历

THEMIS: 面向科研论文造假取证的多模态大模型全面评测基准 2025.08-2025.10

THEMIS: Towards Holistic Evaluation of MLLMs for Scientific Paper Fraud Forensics

- 第一作者, ICLR2026 收录 (Poster)

[基准构建] 构建首个针对学术论文伪造的 MLLM 视觉伪造推理能力评估基准, 涵盖 7 大科研场景、5 种伪造类型 (如 AI 生成, 图文不一致等)、16 种精细化篡改操作, 填补了多模态模型在**专家级视觉推理能力**评测基准的空白。

[任务与指标设计] 设计了单类伪造类型识别与定位、复合篡改操作识别、跨模态不一致性识别与定位三类核心任务; 原创性定义平衡鲁棒性指标 BRI, 填补了该领域量化评估的空白。

[模型局限性分析] 系统评估 16 个主流模型 (如 GPT-5, Qwen-2.5-VL), 揭露了旗舰模型在从伪造识别到伪造定位任务中存在约 60% 的性能骤降。验证了开源模型在特定垂直场景下的性能对等性。

[泛化性与真实性验证] 通过在真实撤稿案例数据集上的交叉验证, 证明合成数据在特征分布与攻击压力上与真实场景具有高度一致性。

[策略优化] 探索并验证了 CoT 与 few-shot 提示词策略对取证任务的影响。针对鉴识逻辑提出, 引导模型进行“异常细节注意-逻辑核查-综合推断”, 实验证明该策略显著优于 Few-shot, 显著提升 MLLM 伪造定位能力 (LLaMa 4 Maverick 平均提升 10.81%)。

AEGIS: 面向 AI 生成内容伪造检测多模态大模型评测基准 2025.04-2026.01

AEGIS: A Holistic Benchmark for Evaluating Forensic Analysis of AI-Generated Academic Images

- 共同第一作者, ACL 2026 收录 (Main Conference)

[数据与基准构建] 构建首个针对学术论文图像 AI 生成伪造取证评估基准, 涵盖 7 大学科学术类别、39 种子类, 利用 25 种前沿生成模型 (混合模型,) 模拟四类核心伪造手段。

[任务与指标设计] 设计了从全局辨别到局部精细定位的四阶取证任务; 提出 NFI (归一化取证指数), 实现跨模型架构的公正性能评估。

[模型能力评测分析] 评测 GPT-5.1、Qwen-2.5-VL 等 25 个模型。实验证明即便顶级模型 (GPT-5.1 综合分 48.80%) 在精细定位上仍有显著瓶颈。

[提示词策略增强方法] 采用 CoT 方法使得 MLLM 在文本伪影识别任务提升至 84.74%。

学术论文图文一致性检测技术研究 ReasoningLab (BUPT) 2025.03-2025.04

针对检测论文工厂图文不一致特征识别提升, 构建 AITCA 数据集 (含 21,103 对图文数据,)

[微调增强] 基于 LLaMA-Factory 框架对 LLaVA-1.5、LLaVA-Next 及 Qwen2-VL 等主流多模态大模型进行 Lora 微调。成功将模型在学术取证任务中的表现提升 20% 以上。