密言

miyan@bupt.edu.cn · (+86) 155-8818-6051 · M1YAN · M1YAN's Blog

研究兴趣：推荐系统安全、大语言模型安全、可信的人工智能、可解释的人工智能……

教育背景

北京邮电大学, 北京 2022 年 9 月 – 至今

在读本科生 未来学院计算机科学与技术专业, 预计 2026 年 6 月毕业

GPA 86.13/100

实习经历

人工智能学院模式识别实验室（PRIS）, 北京邮电大学 2024 年 5 月 – 2025 年 2 月

指导 李思副教授

智能算法安全重点实验室, 中国科学院计算技术研究所 2025 年 5 月 – 至今

指导 曹绮副研究员

项目经历

ReContraster: Making Your Posters Stand Out with Regional Contrast 2024 年 11 月 – 2025 年 3 月

作者 Peixuan Zhang, Zijian Jia, Shuchen Weng, Ziqi Cai, Yan Mi, Si Li, Boxin Shi

关键词 Poster Generation, Image Generation, Automate Graphic Design

主要工作 构建了一个框架，实现含有区域对比画面的海报生成

Goal-Aware Identification and Rectification of Misinformation in Multi-Agent Systems 2025 年 3 月 – 2025 年 5 月，EMNLP'25 under review

作者 Zherui Li, Yan Mi, Zhenhong Zhou, Houcheng Jiang, Guibin Zhang, Kun Wang, Junfeng Fang

关键词 Multi-Agent System based LLM，Misinformation injection attack

主要工作 构建了一个数据集 Misinfotask，该数据集用于评估 Misinformation 注入对 Multi-Agent System 的影响，提出一个通用和适应性的框架 ARGUS 用来防御攻击。

Decoupled Explainable Agent for Recommender System Defense 2025 年 6 月 – 至今

关键词 Agent, Recommender System Defense, Recommender System Safety

主要工作 构想了一个解耦式、可解释的推荐防御智能体，旨在保护推荐系统免受数据注入等恶意攻击。该智能体作为独立的后处理模块，在不侵入推荐系统的情况下，对生成的推荐列表进行实时分析与净化，并为防御行为提供解释。目前工作处于实现和验证阶段。

技能

- 编程语言: Python == C > Shell > Verilog == Matlab
- 工具: Linux, Conda, PyTorch, Git, LaTeX
- 开发: 嵌入式系统设计工程师、静态网站的部署和开发

获奖情况

全国一等奖, 第六届全球校园人工智能算法精英大赛——算法创新赛 2024 年 12 月

北京市一等奖，RAICOM 机器人开发者大赛——机械臂虚实结合挑战赛 2024 年 11 月

最佳性能奖，日本电气通信大学 PBL 国际项目 2024 年 8 月

三好学生，北京邮电大学 2022-2023 学年

优秀团员，共青团北京邮电大学 2023-2024 学年