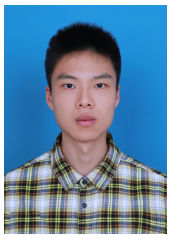
吴胤祺

136-1672-5327 wuyinqi@mail.dlut.edu.cn



教育经历

大连理工大学 - 本科 - 软件工程 2023.09 - 至今

- 前五学期推免排名课程均分 **95.04**，排名 **2/397** （前 **1%**）。
- 主修课程：数据结构与算法（95）；计算机网络（97）；计算机组织与结构（93）；人工智能基础（99）
- 英语水平：国家英语四级为 604 分、六级为 610 分
- 编程水平：CCF-CSP认证考试中获得 **320** 分

获奖情况

奖项荣誉

- 2023-2024 学年国家奖学金、院级学习优秀一等奖学金
- 2024-2025 学年国家奖学金、院级科技创新奖学金

学科竞赛

- 第50届国际大学生程序设计竞赛（ICPC）成都站银牌
- 第50届年国际大学生程序设计竞赛（ICPC）EC Final铜牌

科研经历

RepDAN:Representation Alignment Based Jailbreak Attacks for Large Language Models KDD二作在投

- **研究方向：** AI Safety
- **个人贡献：** 作为主要完成人，主导了 RepDAN 框架的构思，并独立完成了所有代码实现及全流程实验验证。
- **当前工作存在的问题：** 基于梯度的攻击存在高困惑度（易被防御）缺陷，而依赖辅助 LLM 的优化方法因模型自身的安全对齐偏见，极易导致“先肯定后拒绝”的失败模式。
- **核心方法：** 提出了表征对齐新范式，设计了目标引导的单Token损失与轨迹对齐损失，实现了无需辅助LLM的高效离散提示词优化；在GPT-3.5-Turbo上达到了88%的攻击成功率。

AgentChat 科研项目

- **研究方向：** AI Agent
- **个人贡献：** 负责 Agent 核心引擎的代码实现。实现了三层记忆管理模块，构建了基于MCP协议的通用工具调用框架，完成了40余个水文业务接口的标准化封装。
- **当前工作存在的问题：** 原生 Agent 架构在多轮复杂交互中存在严重的“长程遗忘”现象；在执行大跨度历史数据查询时，工具返回的 Token 激增会直接导致上下文窗口溢出，致使推理链中断或系统崩溃。
- **核心方法：** 实现了动态压缩与去重算法通过异步机制自动提取关键结论并裁剪冗余记录。

面向生理辅助任务与情感支持服务的看护机器人系统 大创项目 尚未结项

- **研究方向：** 具身智能、多模态感知
- **个人贡献：** 负责代码实现
- **当前工作存在的问题：** 传统指令解析模型难以处理模糊语义，且感知模块与控制逻辑之间存在严重的耦合，导致扩展性差、无法应对复杂照护需求。
- **核心方法：** 利用提示词工程，通过将多模态情感识别数据转化为结构化上下文，实现了自然语言指令到复杂任务序列的端到端拆解与闭环控制 。

实践经历

- **CPC副组长** 担任大连理工大学创新创业实践中心CPC组副组长。