

林东定 LIN DONGDING

📍 香港 • 📞 (+86) 137 5006 5371 • 📩 22037064r@connect.polyu.hk

GitHub: github.com/DongdingLin • Google Scholar • 求职方向: 大模型算法工程师

个人概述

香港理工大学电子计算学系博士研究生，研究聚焦 **LLM 推理、对话系统、多模态大模型与情境会话推荐**。熟悉从任务定义、数据构建、模型设计到实验评测的完整研发流程，并具备工程落地与迭代优化能力，相关工作发表于 ACM MM、ACL、AAAI、TOIS、TNNLS 等。

教育背景

| | | |
|--------|---------------------------------|----------------------------|
| 香港理工大学 | 电子计算学系博士研究生, GPA 4.18/4.3 | 2022.09 – 至今 • 香港, 中国 |
| 中山大学 | 计算机技术硕士（推免/保研）, GPA 3.9/4.0 | 2017.09 – 2020.07 • 广州, 中国 |
| 中山大学 | 软件工程学士, GPA 3.8/4.0 (排名 37/433) | 2013.09 – 2017.07 • 广州, 中国 |

实习与研究经历

| | |
|----------------------------|--------------|
| 华为香港研究中心 (HKRC) Fermat Lab | 2025.08 – 至今 |
| Research Intern, 优秀实习生 | 香港 |

- 无线在线训练数据筛选 (基站资源优化)**: 面向在线训练场景, 设计本地数据筛选与分布外 (OOD) 过滤策略; 结合蒸馏式判别与 GPU-KMeans 聚类, 在基本不损性能的前提下将训练算力降低 **4x**, 性能衰减控制在 **3%**, 并提交技术报告和相应的代码实现, 形成可复用配置与日志闭环, 支持快速回滚与复现实验。
- 数学题数据自动泛化 + 形式化校验 (Math500/竞赛/K12)**: 搭建自动流水线: 变量抽取 → 参数泛化 → Mathematica 求解与一致性校验, 并输出结构化标注以支持多题型扩展, 支持批量生成; 将 Math500 泛化率从 **42.74%** 提升至 **58.6%**, 新题准确率从 **87.4%** 提升至 **99%**; 竞赛题泛化率从 **7.8%** 提升至 **63.95%**; 在 K12 多题型场景引入多路泛化策略, 将整体泛化率提升至 **42%**; 核心能力封装为 **SDK** 交付落地。
- 同分布 (IID) 对话数据筛选**: 为模拟真实用户提问并用于新模型效果评测, 构建同分布且高质量的对话数据; 以句向量检索与聚类结合规则清洗, 剔除 ASR 噪声、拒答与离群样本, 沉淀高价值 IID 数据。

| | |
|-------------------------------------------|--------------|
| 香港理工大学 NLP Group | 2022.12 – 至今 |
| PhD Research, Supervisor: Prof. Wenjie Li | 香港 |

- 情境会话推荐 (SCR) 建模与评测**: 围绕场景、上下文与用户偏好, 梳理交互链路并定义任务设定与评测指标; 实现数据处理、训练与评测的端到端流水线, 支撑系列论文与复现实验。
- SiPeR: 场景迁移 + 贝叶斯偏好推断**: 面向 SCR 的“去哪/选什么 (Where/What)”决策, 同时建模场景迁移与目标场景预测, 估计当前场景是否满足需求并在必要时引导迁移到更合适场景; 利用多模态大模型 likelihood 做贝叶斯逆推断以刻画隐式偏好, 在 SIMMC 2.1 与 SCREEN 上平均提升 **10.9%/10.6%**。
- Re2A: Rubric 驱动的偏好推理与双对齐生成**: 将 SCR 设计为 reason-then-align 流程: 用自动 Rubrics 提供多维奖励信号, 实现无需人工标注的显式偏好状态推理; 再用 Preference-conditioned DPO 同时对齐“偏好满足 + 场景一致”, 在 SIMMC 2.1 上将幻觉率从 **25.4%** 降至 **5.2%**、弱匹配率从 **22.6%** 降至 **7.4%**。

| | |
|-------------------------------------------------|-------------------|
| 香港理工大学 NLP Group | 2021.07 – 2022.12 |
| Research Assistant, Supervisor: Prof. Wenjie Li | 香港 |

- 目标导向对话规划与生成 (框架工程化)**: 围绕“目标约束、对话状态与用户偏好”设计对话规划与生成方案, 引入 look-ahead/反馈修正等策略提升多轮可控性与目标达成; 并将数据预处理、状态/意图建模、规划器与生成器模块化封装, 统一训练/推理接口与评测脚本, 支持多策略快速对比与复现实验。

- 机器阅读理解与多跳推理（层级传播 + 记忆流）：面向长文档证据聚合与跨句关系建模，提出分层信息传递与记忆更新机制增强证据追踪与信息保真，搭建训练/评测流水线并开展系统性消融与误差分析，沉淀可复现实验代码与配置。

代表论文 (Selected)

- Dongding Lin, Jian Wang, Chak Tou Leong, Wenjie Li. SCREEN: A Benchmark for Situated Conversational Recommendation. *ACM MM 2024* (CCF-A).
- Jian Wang, Dongding Lin, Wenjie Li. Target-constrained Bidirectional Planning for Generation of Target-oriented Proactive Dialogue. *TOIS 2024* (CCF-A).
- Jian Wang, Chak Tou Leong, Jiashuo Wang, Dongding Lin, Wenjie Li, Xiao-Yong Wei. Instruct once, chat consistently in multiple rounds. *ACL 2024* (CCF-A).
- Jian Wang, Yi Cheng, Dongding Lin, et al. Target-oriented proactive dialogue systems with personalization. *EMNLP 2023* (CCF-B).
- Dongding Lin*, Jian Wang*, Wenjie Li. COLA: Improving Conversational Recommender Systems by Collaborative Augmentation. *AAAI 2023* (CCF-A).

学术服务与教学

- 审稿服务：ACL Rolling Review (ARR)、ACL、EMNLP、ACM MM。
- 助教课程：自然语言处理 (2024/25 S2, 2023/24 S2)、数据结构与数据库系统 (2024/25 S1)、移动计算 (2023/24 S1)。

技术栈

编程语言 Python, C/C++, Java, JavaScript, SQL, MATLAB

深度学习工具 PyTorch, TensorFlow, Hugging Face, Scikit-learn

兴趣研究方向 LLM 训练与推理、对话系统、会话推荐、多模态理解

荣誉奖项

- 百度 2021 语言与智能技术竞赛：4/750
- Kaggle 材料质量预测竞赛：4/119
- 中山大学优秀毕业生（前 3%）、优秀毕业论文（前 3%）
- 中山大学本科与研究生阶段多次一/二/三等奖学金