

胡冲

(+86)19946254349 | chong034@e.ntu.edu.sg

教育经历

同济大学 | 软件工程, 软件学院 | 工学学士 2018.09—2022.06

相关课程: 操作系统、大型数据库、软件架构与设计模式、算法设计与分析。获得“华为杯”企业与 CEO 挑战赛 (同济大学三等奖)、中国大学生“互联网+”创新创业大赛 (同济大学三等奖)。

南洋理工大学 | 人工智能, 相关学院 | 理学硕士 2024.08—2026.01 (预计)

相关课程: 深度学习及应用、人工智能数学、机器学习, 文本数据管理与处理。

技术能力

- 语言: 具备良好的编程能力, 无特定语言限制。常用 Python、C++; 熟悉 Java;
- 工作流: 熟练使用 Linux、Git, 了解 Shell 脚本。
- 其他: Transformer, T5 model, KG(neo4j, scrapy, NER 识别, 本体建模), RAG(GraphRag, langchain, ElasticSearch, HyDE)

工作经历

新加坡科技研究局 (A*STAR) | 前沿人工智能研究中心实习生 2024.11—至今

- 致力于增强人工智能的可解释性, 在强化学习系统中集成大语言模型。
- 构建基于多变量状态变量的数学模型, 借助大语言模型和环境模型生成反事实状态空间, 定义相关评估函数以评估状态差异和可供性。

中国银联数据服务有限公司 | 软件开发中心软件工程师 2022.09—至今

- 协助进行下一代核心发卡系统的开发与迁移工作, 使用分布式数据库 GoldenDB。
- 参与现有核心系统的多维账户和个性化定价子系统开发。
- 在银行交易数据库并行处理的批处理优化中发挥关键作用, 处理效率提升 25 %。
- 与多团队协作, 为业务模块转型和系统后端功能提供解决方案, 支持消费贷款核心业务产品开发。荣获“奋斗杯”上海青年金融业务创新大赛 (上海二等奖)。

项目经历

银行业大语言模型辅助信息提取优化方案 | 团队负责人 2023.08—2023.12

- 带领 4 人团队, 基于 ChatGLM3-6b 构建企业级知识检索系统。
- 设计并实现基于大语言模型的两阶段信息提取系统, 优化金融场景少样本学习能力, 降低上下文内存成本 30%。

云原生 Kubernetes 运维知识图谱¹ | 开发者 2021.03—2021.08

- 参与同济大学 Xlab 团队和华为联合开发的 Kubernetes 集群运维知识图谱建模。
- 用 Python 和 Flask 对 Kubernetes 集群进行动态本体建模, 构建 Kubernetes 状态模型, 并用 Prometheus 监控节点实现实时运维。

济星云 (Ji Nebula)-为同济学生服务的门户系统 (与同济信息办联合开发) | 技术团队核心开发者 2022.01—2022.09

- 参与开发“济星云”微信小程序以及 web 端管理中台。深度参与后端架构搭建, 采用高可扩展性的微服务架构设计, 将系统拆分为多个独立的服务单元。
- 使用基于 Spring Boot、MVC 的框架和 MyBatis 实现了多个功能模块: 年度账单, 校园猫咪图鉴, 用户反馈等。采用阿里云对象存储服务 (OSS) 实现图片的分布式存储和快速访问。CI/CD 方面, 依托阿里云效平台达成高效 DevOps 流程。
- 定期维护服务, 累计注册用户自 33000 人扩展至超 45300 人。

研究经历

使用 T5 模型在 Lean 语言中的自动化推理证明 | 撰写中, 第二作者 2024.11—至今

- 基于 98734 条 theorem、217776 种 tactics 和 129243 个 premise, 优化损失函数, 使检索模型的 Recall@10 提升 3%。
- 使用 vLLM 和 DeepSpeed 框架分别优化推理和训练, 配置相关策略降低 GPU 内存使用并提高计算效率, 集成 WandbLogger 进行性能监控和日志记录。

¹ 下划线表示超链接

- 利用矩阵计算和并行处理优化图距离计算，提高 premise 比较效率。

信息熵驱动的 Inference Scaling 算法增强的 GraphRAG 构建 | *ACL* 投稿中，共同第一作者 2025.01—2025.03

- 开发 ESIRAG，通过正交多样化提示、熵引导的自适应扩展和信息增益控制终止，优化知识图谱构建，使 Qwen-plus 等大、小模型在多个 benchmark 上准确率提升最多超 10%。
- 开展知识图谱和本体信息论研究，设计优化熵和粒度指标，基于信息增益确定权重计算以优化熵，结合莱顿社区检测算法构建粒度指标。

个人总结

- 本人积极向上，学习能力强，具备良好的团队协作和沟通能力。
- 英语水平良好，雅思成绩 7.5 分（听力 8 分，阅读 9 分，写作 6.5 分，口语 6.5 分），能够满足工作中的英语交流需求。
- 拥有丰富的软件开发、项目实践和研究经验，持续关注人工智能和软件开发领域的技术发展动态，如 Deepseek-r1, Qwen32B, deepresearch 等。