

李智诚

手机: (+86) 18621669483 | 邮箱: zhichenglee97@gmail.com | 主页: zhicheng0326 | 领英: Zhicheng Lee

教育背景

清华大学, 计算机科学与技术系, 硕士 2023.09 - 2025.06
中国政府奖学金, GPA: 3.80 / 4.00
上海交通大学, 自动化, 本科 2017.09 - 2021.06
上海市政府三等奖学金, GPA: 3.44 / 4.3, 百分制: 83.73

核心能力

- Agent / Harness**: 具备基于 Claude Agent SDK、pi 扩展 Agent 应用的经验, 覆盖多 Agent 编排、工具接入、会话与上下文管理、执行反馈闭环。
- Agentic Engineering**: 熟练使用 AI Coding 工具, 结合 grill-me 驱动需求澄清、CLAUDE.md 工程规范与约束、Skills 和 git worktree 并行开发; 要求 Agent 为功能改动编写测试脚本, 进行自动化测试与回归保护。
- 算法基础**: 熟悉 Python, 具备训练数据构建、SFT 与模型评测经验; 主导 ReaRAG、DiaKoP 研究, 硕士期间产出 2 篇共同一作论文 (IJCNN、CIKM Demo)。

工作经历

智谱 AI, AI 应用工程师 2025.08 - 至今
• 主导企业 Agent 项目的方案与交付, 覆盖 VC 投研、Coding Agent (某车企 ASPICE 标准开发)、多模态生成 (某车企座舱娱乐萌宠视频) 等方向。

智谱 AI, 大模型算法实习生 2024.07 - 2025.05
• **Agentic RAG**: 构建长文本 vs RAG 评测框架, 对比长文本模型 (GLM-9B-128k) 与 RAG, 针对 HotpotQA / NQ 干扰项不足扩展数据集到 50k+ token; 构造训练数据微调推理增强 RAG 模型, 在多跳复杂问答中引入反思与多次检索, 形成 ReaRAG 共同一作论文。

Wise AI Malaysia, AI 工程师 2021.11 - 2023.08
• 主导基于 ResNet 的人脸活体检测研发, 覆盖数据、训练与评测; 通过第三方攻击测试, 达到项目要求的 ISO-30107 Level 1 标准。

AGENT & HARNESS 项目经历

Venture Buddy (VC 智能投研平台)

主导技术方案与 Agent Harness 设计; 产品构思至客户 POC 交付关键负责人

- 应用实现**: 面向某 VC 基金, 基于 Claude Agent SDK 构建投研 Agent 平台, 完成前后端开发, 支持连续对话、工具调用和研究产物展示。
- MCP 工具集成**: 接入搜索、知识库及结构化研究资料读写能力, 按会话装配 MCP 工具与插件, 为 Agent 提供研究所需的数据和操作接口。
- Skills 与投研流程**: 与领域专家深入沟通, 梳理行业扫描、持续跟踪的研究方法与执行约束, 结合专家反馈迭代, 沉淀为可复用的 Skills / Plugins。

某车企 ASPICE Coding Agent (车机单元测试生成)

项目主导; 基于 pi 扩展多 Agent 编排与真实构建环境的反馈闭环

- 任务编排**: 面向 Kotlin 代码库, 由 Orchestrator 协调代码探索、测试编写与评审角色; 探索相关模块和依赖, 按需提供代码、测试约定与工具能力。
- 执行闭环**: 通过项目 CLI、Skills 与跨服务器脚本接入 Gradle 测试和 Kover 覆盖率检查; 将编译、运行失败反馈给相应 Agent 修复并重新验证。
- 上下文与复用**: 分离实现与评审上下文, 以测试执行结果作为反馈; 通过 pi extension 按场景创建 Agent Team, 支持三个客户场景共用生成机制。
- 结果**: 在车机端 Kotlin 业务模块中验证自动化单元测试生成能力, 行覆盖率由 **28.0% 提升至 86.8%**, 分支覆盖率由 **22.3% 提升至 71.3%**; 完成客户 POC 验收。

某车企座舱娱乐 AI 萌宠视频生成（多模态 + 异步任务系统）

- **业务与架构**：为客户 App 提供“上传宠物照片 → 生成动作视频”能力，采用 Redis 队列、Worker 消费与 PostgreSQL 持久化，处理视频生成耗时任务。
- **任务流转**：Worker 轮询模型任务状态，完成后上传对象存储、更新 DB 并同步 Redis 缓存；客户端通过 task_id 缓存优先查询，降低数据库查询压力。
- **多模态 pipeline**：实现 Image→Video 路由，进行跨模型选型与 Prompt 调优，使用 VLM 做输出审核。
- **交付结果**：客户认可生成效果，签收 POC 验收确认书。

算法研究

ReaRAG（推理增强 RAG 的数据构建与 SFT）

共同第一作者；主导数据构建、SFT 训练与评测

- **数据构建**：基于 MuSiQue、HotpotQA、NQ 构建种子问题，用 QwQ-32B-Preview 生成模型推理与检索轨迹，经筛选得到约 **2 万条** 训练数据。
- **模型 SFT**：对 GLM-4-9B 进行 SFT，混合轨迹与通用数据，仅对推理和动作对应的 token 计算损失，强化模型多步检索与回答能力。
- **评测与分析**：在 6 项问答基准上采用 LLM-as-a-Judge 评估答案正确性，结合推理能力消融、检索步数与失败案例，分析模型效果及效率取舍。

学术论文

ReaRAG: Knowledge-guided Reasoning Enhances Factuality of Large Reasoning Models with Iterative Retrieval Augmented Generation

Jinxin Liu*, **Zhicheng Lee***, Shulin Cao, Jiajie Zhang, Weichuan Liu, Xiaoyin Che, Lei Hou, Juanzi Li
International Joint Conference on Neural Networks (IJCNN), 2026

AtomR: Atomic Operator-Empowered Large Language Models for Heterogeneous Knowledge Reasoning

Amy Xin, Jinxin Liu, Zijun Yao, **Zhicheng Lee**, Shulin Cao, Lei Hou, Juanzi Li
31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), Vol. 2, 2025

DiaKoP: Dialogue-based Knowledge-oriented Programming for Neural-symbolic Knowledge Base Question Answering

Zhicheng Lee*, Zhidian Huang*, Zijun Yao, Jinxin Liu, Amy Xin, Lei Hou, Juanzi Li
Conference on Information and Knowledge Management (CIKM demo), 2024

Deep Semi-Supervised Learning via Dynamic Anchor Graph Embedding Learning

Zihao Wang, Enmei Tu, **Zhicheng Lee**
International Joint Conference on Neural Networks (IJCNN), 2021