陈乐 CHEN LE

年龄 23 | 最高学历 硕士 · QS 18 （2026.11 毕业） | 政治面貌 中共党员
电话 17703750681 | 邮箱地址 kinsleylchen@outlook.com | 英语水平 专业八级

教育经历

香港中文大学 · QS 18 **语言学** **硕士** 2025.09 – 2026.11

主修课程： 语音学与音系学、句法学与语义学、语言习得、语料库语言学

大连外国语大学 **汉语国际教育** **本科** 2021.09 – 2025.06

荣誉奖项： 国家奖学金 · 校优秀毕业生 · 社会实践先进个人 · 优秀团干部 · 校奖学金

实习经历

字节跳动 · 模型评测 · 豆包办公 2026.05 – 至今

- **用户场景与 Agent 能力拆解**：面向豆包办公表格、文档等真实用户任务，拆解选区编辑、数据处理、图表生成等场景下的核心诉求，主导构建多维度评测题库体系（选区编辑、金融、错题本、难题、图表等），搭建 **AI 初筛→AAJ 精筛→GT 互检→Rubrics 融合→双 Judge 仲裁 Pipeline**，使被测模型与优质模型稳定拉开 **10+分差**，针对性找出劣势点，支撑 Agent 能力规划与版本迭代。
- **Agent 产品体验优化**：围绕**用户需求是否被正确理解并完整交付**分析 Agent 全链路，通过 **Skill 召回与 trace 下钻**定位需求理解偏差、工具未调用、产物未落地等问题，推动**主 skill 召回率由 78%→87%**，reference skill **提升 14%**，并沉淀典型 badcase 供模型迭代。
- **自动评测方案迭代**：针对 LLMAJ 在复杂 Agent 产物评测中存在的证据漏读、抽样判断及文件级核验不足，通过**裁判审计**定位方案能力边界，推动**评测链路由 LLMAJ 迭代至 AAJ**，将人机一致率提升至 80%+/局部维度最高 99%，提升复杂 Agent 任务效果判断的可靠性。
- **竞品分析与评测工具化**：搭建 Kimi、Claude、WorkBuddy 等**竞品横评基线**，定位公式/图表交付、复杂指令遵循、视觉美化等核心体验差距并输出优化建议；针对评测过程中的数据清洗、异常 case 重跑与结果校验等重复环节，开发多条 Skill + Python **脚本实现流程自动化**，提升多模型效果分析与问题定位效率。

传音控股 · AI 产品运营 · 灵动岛 2026.03 – 2026.05

- **AI 场景与能力设计：**围绕灵动岛航班、日程等高频场景，拆解用户消息从**意图识别→关键实体抽取→场景触发**的核心链路；针对多语言环境下的时间、地点及多意图表达，设计 11 类事件×10+时间格式的识别规则与数据体系，支撑灵动岛场景识别能力优化。
- **产品效果分析与体验优化：**搭建意图级与实体级效果评测体系，基于 F1 指标分析模型效果，拆解航班、日程场景下的意图识别与实体抽取表现；通过 badcase 分析定位边界模糊、多意图冲突、时间表达泛化不足等核心问题并优化，推动**关键实体 F1 提升 23~36%**。
- **优化内部工具**：针对自研 NER 工具在模型结果提取、对比及人工处理流程中的效率问题，优化前端功能与交互，支持 5 个模型标注结果统一提取与对比，使单条样本处理耗时下降 60%，提升数据分析与模型迭代效率。

虾皮（Shopee） · 产品运营 · 卖家学习中心 2025.12 – 2026.03

- **用户洞察与需求分析**：基于平台、企微、内部工单等渠道分析覆盖 **8000+名卖家**的真实反馈，围绕功能使用、物流履约、平台规则、活动搭建等问题进行**建立分类与优先级机制**，提炼高频痛点并形成 FAQ、操作路径及产品优化建议，推动问题闭环。
- **推动功能上线与平台政策落地**：参与 Shopee 标准产品、包裹出口查询等**功能上新迭代**，以及节日履约/物流豁免政策的卖家侧落地，从卖家使用场景出发梳理功能说明、操作路径、异常场景与客服口径；上线后**相关使用咨询量下降 33%**，降低用户理解与使用成本。
- **业务策略与活动搭建**：结合箱包、服饰等类目需求整理选品指南、关键词库与运营素材，支持卖家供给优化和平台品类结构调整；同时围绕新手卖家活动搭建提供店铺装修模板、活动素材与配置指引，降低活动配置与平台规则理解门槛。

项目经历

求职申请管理看板 独立产品设计 *yu9wphwd1xl6.space.minimaxi.com*

- 需求洞察与产品定义：从多平台求职过程中信息分散的实际痛点出发，梳理求职全流程中的信息管理需求，并将其拆解为核心模块。
- 完成产品结构与流程设计，将分散在不同招聘渠道的申请信息统一组织，并通过状态与进度可视化降低信息遗漏和重复维护成本。

基于语料库与大语言模型的汉语话语理解与语序偏好研究 硕士论文 语料库 + 大语言模型 研究

技能： Benchmark 构建 | LLM 评测 | 统计建模 | Python

- **研究问题与数据体系设计**：围绕 LLM 在复杂中文语境下是否过度依赖表层语言模式的问题，基于真实中文语料构建**宾语话题化评测集**，覆盖非典型语序、跨句指代、话语距离等复杂语境因素；结合人工标注与**一致性检验**（Cohen's κ = 0.84）沉淀 210 条样本。
- **LLM 能力分析与问题归因**：借助**脚本+LLM API** 完成语料筛选、语境抽取、自动化标注及**多模型语序判断实验**，通过**逻辑回归**拆解影响模型判断的关键因素，发现模型在跨句语境约束下稳定性不足，为中文大语言模型 Benchmark 设计和难例构造提供参考。

核心能力

聚焦 AI 应用与 Agent 产品方向，具备大模型评测、用户场景分析与产品运营复合经历。能够从真实用户任务出发拆解需求与 AI 能力，通过竞品分析、Agent trace 与数据评测定位产品体验问题，并结合模型能力边界推动优化方案与效果验证。熟悉 LLM/Agent、Prompt、Skills、Benchmark 与自动化评测，可连接用户需求、产品体验与模型能力，推动 AI 能力从“可用”走向“好用”。

用户场景分析 | 需求与能力拆解 | 竞品分析 | Badcase 归因 | Python | Benchmark | LLM/Agent-as-a-Judge | Agent/Workflow