



教育背景

- 北京大学-博士 前沿交叉学科研究院 - 区域与国别研究院 全日制 2022.09 - 2027.06
- GPA: 3.7/4.0 | 获北京大学三好学生、优秀科研奖等
- 中山大学-本科 国际关系学院 - 国际政治系 全日制 2018.09 - 2022.06
- GPA: 3.9/4.0 | 获教育部国家奖学金、中山大学优秀共青团员、优秀学生一等奖学金、王老吉奖学金等

实习经历

- 北京大学区域与国别研究院-战略研究员 2022.01 - 至今
- AI辅助战略研究: 搭建覆盖信息检索、信息筛选、事实核验与报告生成的 AI Research 工作流, 缩短复杂议题的研究响应周期; 独立完成 33 篇研究报告报送, 获中宣部等部委采纳1篇; 发表中英文媒体评论9篇, 形成公共政策影响力。公派赴印度尼西亚大学访学 6 个月, 通过实地调研与企业访谈研究中印尼产业合作, 在真实海外场景中积累用户需求洞察与跨文化信息获取经验。
- 广西壮族自治区外事办公室-综合协调处实习生 2025.08 - 2025.09
- 信息分析与决策支持: 围绕中国—东盟 AI 合作动态与政策趋势, 主笔完成外事内参、领导人谈话参考及调研报告等 7 份决策信息产品, 将分散政策、产业及区域信息拆解整理为可直接服务决策的结构化结论。深度参与第22届中国—东盟博览会统筹工作, 协同多方推进材料准备与活动执行; 将区域国别研究成果转化为面向外办青年干部的专题培训内容, 并主讲“区域国别研究论坛”。

项目经历

- 全球商业洞察 | AI搜索研究 Agent — 独立产品设计与开发 2025.12 - 2026.08
- 需求洞察与竞品分析: 基于6个月海外田野及30+出海企业负责人访谈, 归纳企业国际化过程中在市场筛选、政策判断及风险识别等方面的共性需求; 测试通用 AI Search / Deep Research 工具后发现, 其在发展中国家的本地信源覆盖、权威资料整合与国别差异识别仍有不足, 据此将产品定位为面向企业“一国一策”决策的出海战略研究 Agent。
 - 战略情报体系设计: 搭建覆盖195个国家、2,367个分层信源的 Market Pack; 结合外交部、商务部等311份官方资料构建 RAG 知识库, 形成本地信息获取、权威资料校验与多源证据交叉验证的出海战略研究底座。
 - Agent设计与效果验证: 设计 Harness + multi-agent 架构, 建立从意图识别、问题拆解、检索规划、证据组织到报告写作、引用核验的完整研究链路; 引入“先定义决策问题、再确认研究方向”的交互机制, 由模型自主判断是否需要补充提问, 用户确认后再启动联网检索。建立基于测试集与 Bad Case 的评测机制, 经5轮迭代, Agent综合评分 2.94→3.33 (+13%), 完成部署上线。
- 知香 | 知识问答与商品导购 Agent — 独立产品设计与开发 2026.07 - 2026.08
- 需求洞察与产品定位: 访谈20+电商消费者, 识别香水消费中知识门槛高、线上难试香、知识与购物未打通闭环等核心痛点, 将需求拆解为自然语言找香、多轮澄清、真实商品推荐与偏好记忆等关键场景; 据此明确“知识图谱负责理解、真实商品源负责供给”的产品边界, 定位为兼顾香水知识问答与购买决策的导购 Agent。
 - 知识与推荐体系设计: 构建覆盖1,971款香水、504种原料的香水知识图谱, 由 LLM 完成用户意图识别与需求理解, 结合知识图谱检索、多轮澄清及真实商品源, 实现从“用户模糊需求 → 香水知识解释 → 个性化推荐 → 真实商品推荐”的完整导购链路。
 - Agent设计与效果验证: 建立基础回归、改写外推及 Bad Case 专项测试集, 围绕用户问题改写、召回策略与提示词持续迭代; 经3轮优化, 综合评分2.26→3.14/5 (+39%)、平均维度分2.60→3.52/5 (+35%)。独立完成H5产品上线, 实现流式对话、商品卡片、历史会话、用户反馈与运营监控, 建立从需求识别到部署上线的0→1闭环。
- 国际气候谈判数据分析与 LLM 标注项目 — 项目负责人 2024.01 - 2025.05
- 问题拆解与数据验证: 将“重大国际事件是否真正改变国家合作行为”这一难于观测的问题, 拆解为可量化指标, 基于159国长期互动数据使用 R 建模构建反事实框架, 并采用贝叶斯合成控制法、安慰剂检验及替代指标等方法识别事件净效应, 区分事件影响与既有趋势, 形成基于数据证据的判断, 研究成果入选社科院论坛与清华大学国际关系共同体年会。
 - 问题发现与 LLM 标注方案设计: 研究中发现现有数据库覆盖有限, 而人工扩充面临成本高、更新慢与一致性不足等痛点, 因而探索利用 LLM 扩充互动数据。作为前期数据分析负责人, 向技术团队梳理原始数据类型、分布特征、标注逻辑与典型边界问题, 将研究需求转化为可执行的 LLM 标注方案, 并围绕“模型能否替代部分人工标注”明确模型可用性的评估框架, 从识别准确性、人工一致性、事实幻觉、边界识别及典型 Bad Case 等维度定义验收标准, 协同北大王选计算机研究所推进方案设计与效果验证。
 - 评测与规模化验证: 负责统筹10名评测员建立标注与分歧复核机制, 围绕实体识别错误、关系提取错误、议题边界不清等系统性偏差建立LLM标注错误清单, 通过与技术团队的多轮沟通持续优化, 实现模型Precision 67.5%→74.0%、Recall89.2%→90.0%, 完成45,051条数据的规模验证, 合作论文入选2025 Asian PolMeth Conference。

相关技能

- AI 产品能力: AI Search / Deep Research Agent 产品设计, 用户研究与需求拆解, RAG与联网检索, 模型评测, 数据驱动迭代
- AI Native能力: 熟练使用Codex / Workbuddy 等AI Coding工具, 能独立完成产品设计、数据分析、评测脚本及线上部署