



杨凯隆

硕士研究生 | 2027届应届毕业生

17386097253 | kyle0324xj@gmail.com | 北京-上海-深圳 | auratracer.kyleforge.com

教育背景

 西安交通大学	新闻与传播	硕士	2024.09-2027.06
推荐免试攻读硕士、二等奖学金；研究计算传播、社交媒体、用户行为与新媒体运营			
 湖北大学	大数据管理与应用	本科	2020.09-2024.06
专业前5%、国家奖学金、优秀本科毕业论文、三好学生标兵；主修数据挖掘、机器学习与时间序列分析			

实习经历

 腾讯云计算	大模型数据运营实习生	媒体AI产品中心	2026.04-2026.08
--	------------	----------	-----------------

虚拟试穿训练数据生产

背景与目标：虚拟试穿模型缺少成体系训练数据，商品图散落在10余个时装与内衣品牌网站。需要自动化采集原始图库，用模型筛出可用样本，再从模特图中分割衣物作为换装条件，试穿才不受人物姿态与皮肤干扰。目标建成可复用的三元组与四元组生产线

我的职责：独立负责采集、智能筛选与SAM3分割三条线，并衔接换装生成与交付对账

自动化数据采集：站点结构各异且部分有强风控，按站点在API、HTTP、Playwright与浏览器代拉之间切换；强风控站先枚举产品清单、媒体走直链下载，配合断点续跑、限流冷却与自动去重，缺失与失败分类对账。累计入库**3.1万余款商品、图视频逾11万份**

智能模型筛选：原始图库混杂模特图、平铺图、细节图与无关图。SGLang多卡部署Qwen3.8-27B并叠加投机解码，吞吐较llama.cpp约**2.6倍**；商品多图同屏一次请求，**全库请求量由约16.2万次降至1.3万次**。提示词限定输出互斥枚举的可核实验事实：是否模特图、款式类目、正侧背视角、纯色或花纹各附视觉判据，禁止模型做可用性主观判断；程序规则统一裁决、不采信自评。关闭思考经评测一致率**90%**、单张约**11秒**，正面图可用率**98%**

SAM3分割优化换装：时装站自带平铺图可直接用，内衣站只有模特图，需先提取衣物作为换装条件，否则人物姿态与皮肤会干扰试穿。SAM3按上装、下装、全身分别调参，经盲评冻结生产配置；同款仅对正面图分割、侧背复用，推理量减近六成，单张**2.6秒**

指标与效果：三元组与四元组合计交付**近5万组**，组级重排不重复计数，外部整理与强风控站点数据另行交付。抽检发现错分与颜色不匹配后做两轮全量重申，产出换装与试穿图**近9万张**。交付前逐张校验画布一致、零缺失，数据支撑三类LoRA训练与部署验证

 中视前卫	媒体监测实习生	媒体战略研究中心	2024.11-至今 项目制兼职
---	---------	----------	---------------------

总台智能舆情监测系统

背景与目标：面向总台值班场景的日常舆情监测上报需求，值班员需从大量稿件中快速定位差错与舆情风险。独立开发覆盖采集、审核、数仓、看板与自动交付的全链路系统，以AI分级审核辅助人工，提升监测效率与上报质量

我的职责：独立负责系统全链路设计与开发，并按业务口径持续迭代

数据链路：以Wisky采集微博、百度、头条热榜及重点账号稿件，命中后回各平台补爬正文；评论覆盖B站、抖音、百家号，HTTP与Playwright处理登录、分页和字段差异，热榜快照**1,145条**追踪重复上榜话题，数据统一入库以SQL管理、按日累加，值班日约**2,000条**稿件

模型审核：经规则去噪与模型路由调度豆包Seed 2.1、DeepSeek-V4等做初审与帖子级精判，全量稿件先由AI把关、高危候选才转人工复核；提示词注入日期与近期事件，只提取原文可核对差错、拦截知识截止误报，用量约**200万tokens/值班日**。值班员从逐条浏览约**1,000条**稿件转为仅复核约**25条**高危候选、约4条进入速报，评论负面约29条

数据建模：自建**17表**SQLite数仓并按ODS-DWD-DWS-ADS四层建模，汇聚6类数据源完成清洗、脱敏与增量入库；以窗口函数环比、漏斗拆解与台账多表关联完成业务SQL，指标字典统一口径与分母、漏斗双口径如实并列，标注与上报实时写仓、去重读仓反哺初审

系统落地：FastAPI控制中心调度全链路，ECharts搭建漏斗、Token、对账等**7个看板**，Token成本按日、端点、事件三维归因；值班日约每半小时监测，**终态事件成功率达95%**，自动生成XLSX报表、DOCX速报上传云文档，上报及总台反馈条数居部门第一、绩效奖励月均10余次

 深信科创	开源社区运营实习生	市场销售部	2025.03-2025.08
---	-----------	-------	-----------------

CARLA技术问答智能体与知识库

背景与目标：CARLA中文社区的问答散落在CSDN、知乎与GitHub，新人入门门槛高，人工答疑重复且成本高。目标是用问答智能体沉淀可复用知识库，以内容自动化扩大社区触达，推动社区服务产品化落地

我的职责：负责问答智能体搭建与技术内容自动化两条线，覆盖知识库建设、渠道部署、内容生产与效果复盘

问答助手：将CSDN、知乎、GitHub上**1,000+条**问答清洗去重后建成问题—回答知识库，覆盖源码编译、OpenDRIVE、传感器与Apollo/SUMO联合仿真等高频主题；基于Coze、Dify与火山引擎API在微信公众号、小程序、社群等渠道部署，按检索、生成与拒答三级作答，未覆盖问题自动回写知识库持续迭代；上线后日均服务提问约10次、答复采纳率约**90%**，替代约**50%**重复人工答疑

技术内容自动化：每日定时拉取论文与技术动态生成选题，为CARLA中文站及公司公众号发布**技术内容30余篇**；另爬取智联、Boss直聘等平台**176→520个**CARLA相关岗位，经Pandas清洗与sentence-transformers语义聚类提取算法/研发/测试/产品**四类职责群组**，产出薪资-学历-经验交叉分析及《会CARLA的人近9成年薪超20万》等人才调研推文，最高阅读1,000+、带动公众号涨粉50+

 学而思国际	语培新媒体运营实习生	学习事业部	2024.01-2024.05
--	------------	-------	-----------------

新媒体内容运营与运营数据分析

背景与目标：面向GRE备考用户的内容获取与咨询转化需求，负责站外新媒体矩阵的内容生产与数据复盘，用内容与数据两条线支撑引流、转化与留存

我的职责：独立负责内容运营与数据分析两条线，覆盖选题排期、内容产出、多账号运营与策略复盘

内容运营：独立负责GRE业务站外运营与公众号运营排期，围绕考试节点与用户需求统筹热点及学术类推文，累计发布**60+篇**；负责小红书官方账号内容撰写与引流转化并持续运营单词垂类账号，通过选题策划、发布节奏与封面标题的持续对比调优，累计发布**50+篇**、带动粉丝增长**1,000+**，稳定为销售侧导入咨询线索**200+/月**

数据分析：搭建日常运营数据监测体系，追踪内容曝光、互动、引流转化与用户反馈等核心指标，定期输出运营数据报告并向销售顾问与主讲老师同步业务情况；周期性开展竞品分析与活动内容复盘，据此动态调整内容规划与运营策略

项目与研究经历

科技新闻事件追踪引擎

独立开发

2026.08-至今

背景与目标: 调研26个新闻聚合与阅读产品并检索GitHub开源方案, 发现现有工具多聚焦热榜聚合或单篇摘要, 较少覆盖跨中英文信源的事件级聚类与持续追踪, 遂从零开发科技事件追踪引擎, 把同一事件的多源报道自动归并为可溯源的事件页

我的职责: 独立完成需求调研、架构设计、采集管线、模型选型与提示词工程、可视化、测试运维与开源发布

多源采集: 基于YAML领域包管理**152个中英文信源**, 覆盖科技媒体、国际媒体、企业官方与Hacker News/GitHub Trending等热榜; 统一封装RSS、HTML、JSON、sitemap与归档五类采集适配器, 按信源热度实行5-60分钟分级调度

六段式LLM管线: 初筛、聚类、忠实摘要、事件框架、翻译与热度计算; 本地部署Qwen3.5-4B与腾讯混元Hy-MT2, 关闭思考模式后结构化输出稳定可解析、单条约1秒, 翻译0.2-0.4秒/条且在自建评测集上未观察到专名及数字错误, 双模型同机约占5GB显存; 摘要按倒金字塔、不超过300字、只用原文事实且逐句可溯源

跨语言聚类: 用Qwen3-Embedding做跨语言向量匹配, 采用词面规则、向量召回、灰区LLM裁决三级策略; 同事件相似度0.90-0.92、非同事件0.39-0.55, 在100条人工标注事件对上聚类判定准确率**84%**

工程与验收: 仅向渲染层暴露标题、AI摘要、元数据与原文URL, 从架构上避免正文进入静态站; Jinja2生成中英双语静态站与7类SVG可视化, GitHub Actions每30分钟自动重建; **161项pytest**、28项Playwright E2E, 覆盖聚类、构建、i18n与死链检查

指标与效果: 累计处理**9,400+新闻条目**、**3,972个事件**, 均支持中英双语页面; 单轮完整管线 $\leq$ 20分钟; 核心LLM推理全部本地运行, 无外部模型API成本; 已开源 (MIT、中英README、CI与贡献指南), 在线页面 auratracer.kyleforge.com

社交媒体数据挖掘与计算传播研究

第一作者

2024.08-2025.12

背景与目标: 围绕社交媒体传播效果开展两项计算传播研究: 短视频传播归因获CSSCI期刊拟录用, 微博突发事件舆情演化发表于《信息技术与管理应用》

我的职责: 第一作者, 负责研究设计、数据采集、建模实现与论文撰写

短视频传播归因 | 多模态特征 + CatBoost + SHAP

数据与特征: 从79个百万粉丝级县级融媒体抖音号中筛选20个头部账号, 采集2025年6-8月**近2万条**短视频, 剔除带货广告后按互动上四分位筛出**2,014条**头部样本; 以信息生态与多模态理论构建四维框架, 豆包1.6-vision大模型编码文本语义, 人工编码一致性Cohen's Kappa**0.84**, Librosa、OpenCV与PySceneDetect抽取响度、节奏、亮度、色彩与镜头切换等**23维特征**; 传播指数按DCI对数加权, 「大家都在搜」度量平台可见性

建模与结果: 对比五类模型后采用CatBoost, **AUC为0.7788**; SHAP归因显示平台可见性贡献最高**0.847**、听觉特征驱动强于视觉; 经阈值敏感性与稳健性检验, 自然流量与推荐流量下核心要素存在差异. 第一作者论文获CSSCI期刊拟录用

微博突发事件舆情演化 | NLP + 主题模型

演化分析: 爬取2023年全年微博热搜**154,632条**, 结合人工判定、腾讯云NLP「文本分类V2」与StructBERT情感模型筛出**14,841条突发事件**; HanLP加载自建词典(地名/人名/机构名/网络流行词)分词并去停用词, 百度AipNLP按正负概率与置信度加权计算情感得分, Tomotopy LDA以主题一致性 with 困惑度结合scikit-optimize贝叶斯优化(min_cf/min_df/rm_top)识别自然灾害、社会治安、人身安全、国际冲突**4类主题**; 按季度4期计算主题词袋余弦相似度追踪内容演化, NetworkX按度中心性绘制语义共现网络、热力图与桑基图, 从热度、情感、内容三维刻画演化过程; 成果发表于《信息技术与管理应用》

电力需求预测算法挑战赛

参赛选手

2024.07-2024.08

背景与目标: 科大讯飞电力需求预测赛题, 基于训练集**2,877,305行**、测试集58,320行、19类房屋数据预测未来10天用电量

我的职责: 负责特征工程与模型融合方案

时序特征: 按id分组构建10-35天及2-12个月历史平移、1-3阶差分、7-360天滚动均值、极值与方差, 并用OpenFE两阶段自动构造组内排名、对数变换等特征, 经相关性筛选降维

模型融合: 各房屋序列不连续, 采用shuffle的5折K-Fold而非时序切分; 对LightGBM、XGBoost、CatBoost分别训练, Optuna贝叶斯调参并可视化跟踪, 三模型预测以Ridge为第二层做Stacking融合降低单模型偏差

指标与效果: **初赛MAE为215.14567, 排名13/1605、位列前1%**; 获Datawhale AI夏令营二等奖

无糖饮料消费现状与健康发展研究

核心成员

2021.12-2023.04

背景与目标: 省级大创课题, 围绕无糖饮料消费行为开展实证研究并参与全国市场调查与分析大赛

我的职责: 核心成员, 负责问卷调研建模与电商评论挖掘

回购预测: 基于李克特五级量表设计问卷, 经预调查(有效回答率80.15%)确定样本量后以PPS三阶段抽样回收752份有效问卷(线上504、线下245)并完成信效度检验, 用GAN将样本扩充至5倍, 对比Logistic、决策树、SVM与随机森林后选用XGBoost, 网格搜索调参; 特征重要性显示广告宣传、健康属性、联名推广为回购主要因素

电商评论与客群挖掘: 爬取淘宝、京东5,400余条评论, 微调BERT做情感三分类, Biterm短文本聚类出配送品质、产品口味、健康安心三类主题; 对115名未购买用户9项特征做k-modes聚类, 划分年轻新潮型、理智稳重型、从众优享型客群

指标与效果: **验证集准确率90.08%**; 基于特征重要性输出面向品牌的4P营销策略建议; 项目获全国市场调查与分析大赛总决赛二等奖

荣誉与技能

竞赛奖项: 第十二届全国大学生市场调查与分析大赛 (**全国总决赛二等奖**, 2022.05); 第九届中国国际「互联网+」大学生创新创业大赛 (湖北省复赛决赛银奖, 2023.09); 第十三届全国大学生电子商务「创新、创意及创业」挑战赛 (湖北省选拔赛二等奖, 2023.08)

数据与分析: SQL/SQLite 数仓 ODS-DWD-DWS-ADS 建模与漏斗/环比指标; Python/Pandas 清洗、特征工程与Stacking多模型融合 (LightGBM/XGBoost/CatBoost); 时序预测、问卷调研与信效度检验; ECharts与手写SVG可视化; 客群分层与经营分析报表

大模型应用: 本地llama.cpp部署Qwen3.5-4B、混元Hy-MT2、Qwen3-Embedding, SGLang部署Qwen3.8-27B做多模态筛选, SAM3分割; 结构化输出、提示词工程与评测门禁; API调用豆包、DeepSeek、Claude、Gemini与Grok

采集与工程: 多源异构采集、跨语言事件聚类、六段式LLM事件追踪管线; 舆情分层审核与情报速递; GitHub Actions定时重建静态站、FastAPI调度、Docker/Caddy网关

其他经历: 曾任校团委宣传部副部长、学院宣传中心负责人、学院办公室助理、Datawhale鲸英助教, 自营小红书账号**粉丝2,000余人**