殷龙

+(86) 18662513062 ◇ 14 年工作经验 ◇ 大数据架构师/AI 大模型研发工程师

教育

华东师范大学（985,211） *2022.09 - 2025.06*

软件工程（硕士），GPA:3.61/4.0

研究课题：自然语言处理、面向表格文本混合数据的数值推理研究（企业财报理解）

科研成果：一篇 CCF C 类论文、两项专利、两项软著，Multihiertt 数据集 CodaLab 竞赛排行榜第一名

河北工业大学（211） *2006.09 - 2011.06*

计算机科学与技术（本科），GPA:3.1/4.0

法国巴黎高等计算机学院（Supinfo） *2006.09 - 2011.06*

计算机科学与工程（学士）

自我评价

拥有 10 年以上互联网 IT 相关项目及产品开发经验，对车载终端、嵌入式、互联网金融、AI+ 教育领域产品有丰富经验。熟悉 Java、K8s、Hadoop 大数据平台，知识图谱和图数据库、Spark/Scala 及 GraphX Pregel API 和相关的图挖掘算法，熟悉 Python、Pytorch 具备机器学习和深度学习的相关理论基础，熟悉 NLP 相关技术，熟悉大语言模型及其微调（SFT）、llama-factory、vLLM、deepspeed、LangChain、LangGraph、RAG、Ollama、Agent 相关研发。喜欢钻研，对待技术保持归零心态，追求突破。

技能和证书

- 职业资格证书（三级/高级技能）：计算机程序设计员
- TigerGraph GSQL 101 认证
- 计算机技术与软件专业技术资格（软考中级）软件设计师
- 英语六级（521 分）

荣誉和成果

- 分别获得大学第一、二、三学年三等奖学金，第四学年二等奖学金
- 获得 09-10 学年 Supinfo 优秀学生奖
- 获得 2012 年苏州工业园区第三届高技能人才职业技能竞赛“计算机程序员（软件开发）”优胜奖
- 软著：基于预训练语言模型的企业财务报告问答系统
- 软著：基于语言模型的企业财务报告问答数据集处理与标注系统
- 专利：一种基于增量数据计算连通图在股权穿透上应用的优化方法
- 专利：用于电脑的知识图谱管理平台图形用户界面
- 专利：基于图搜索算法的自适应学习方法及计算机学习系统
- 专利：一种面向文本与表格混合数据问答的检索方法、检索系统及应用
- 会议论文：Long Yin, Kai Yin, Hui Zhao, “A Unified Framework for Knowledge-Intensive Numerical Reasoning over Financial Document”, The 19th International Conference on Document Analysis and Recognition (ICDAR 2025), CCF C, Accepted.

工作经历

问向教育科技（上海）有限公司 *2024.09 - 2025.06*

大数据架构师

- 负责主导设计与实施实验室学生画像的宽表增量优化工作，优化后使 DataWorks 日均计算成本从 300 元降至约 2 元。
- 负责主导开发基于 MLflow 的 AI 教练评分结果一致性与稳定性评价框架，使迭代验证周期从原先的 3 天缩短至 2 小时。
- 负责主导问向学生关爱平台 AI 教师助手开发，在关爱案例检索场景中使检索准确率（Hits@3）提升至 0.93。
- 负责大模型备案中涉及到的数据标注平台开发、组织数据标注任务以及模型微调训练。

同盾科技（上海）有限公司 *2021.08 - 2024.08*

架构师/技术经理

- 负责知识图谱平台云图产品的底层技术架构设计，制定兼顾灵活性与通用性的技术架构与选型方案。
- 负责云图、明模、天策、Ark 等产品依赖的大数据组件相关技术调研，新技术探索。主导大数据开发过程的标准化方案制定，提升数据开发效率。
- 主导云图产品底层组件（Hadoop/Spark GraphX pregel API）图挖掘算法性能优化方案研究和实施，优化股权穿透算法的执行效率提升 42 倍，优化连通图相关算法效率提升数十倍。
- 负责图数据库（JanusGraph/Nebula）读写性能优化，数据挖掘算法流程和性能优化。
- 负责优化平安租赁和邮储智慧大脑的百亿级图数据挖掘性能。
- 负责华东区域银行项目技术经理职责，如平安租赁、兴业银行、华瑞银行、银联商务和南银消金的云图、天策等项目实施。

上海精锐教育科技有限公司 *2020.04 - 2021.07*

大数据技术总监

- 负责精锐数据中台搭建、指标梳理、底层数据模型设计和重构，数据开发工具相关的技术选型。
- 带领团队负责精锐数据门户的设计和开发工作，复杂业务逻辑场景下的开发模式设计。
- 负责基于 SimBERT 的拍照搜题 App 的技术调研、设计和开发维护的相关工作。
- 负责集团日常报表的开发与维护工作。

上海乂学教育科技有限公司（松鼠 AI） *2018.04 - 2020.04*

大数据架构师

- 带领研发团队开发并上线教研数据分析平台 xdatam，负责教学数据采集、教研报表的设计与开发。
- 设计基于 Hadoop、GreenPlum 和 Kylin 的大数据基础服务架构、数仓分层设计、拉链表设计、增量数据指标设计和技术方案落地。
- 主导使用 Spark Stream 实时算法处理学生水平包括知识点掌握率、攻克率、学习效率等指标的设计和开发工作。
- 主导教学系统中智适应引擎的算法调研与开发，研究并工程化实现基于项目反应理论（IRT）的学生能力评估模型，用于支持做题过程中的个性化推荐。
- 主导 BKT 模型用于 stop policy 检测学生学习终止点的参数训练和验证工作。
- 负责知识图谱平台的架构设计、功能设计，主导图谱平台微服务设计和开发工作。

上海旺资融资租赁有限公司 *2017.03 - 2018.04*

系统架构师

- 带领研发团队 12 人开发并上线汽车金融融资租赁平台“予见”，显著提升了审批效率与业务响应速度，风控规则迭代周期从 1 周缩短至 1 天。
- 研发国内首个汽车融资租赁平台智能风控系统 Credit Matrix。该智能风控系统采用机器审核加人工审核的模式，审批效率提升 40%，风险识别准确率提升 30%。
- 主导设计和开发数据平台，图数据库存储，关联人网络的领域建模和反欺诈模型设计。
- 负责旺资车贷平台的全栈架构设计与后端核心开发。

苏州思必驰信息科技有限公司 *2011.07 - 2017.02*

高级软件开发工程师

- 负责并主导研发国内首款智能语音导航车载 HUD 终端“车萝卜”产品及其硬件核心驱动。
- 主导开发并维护京东商品数据抓取项目，完成分布式爬虫设计、KV 缓存设计以及商品关键词检索与统计功能。
- 主导研发国内首个英语口语评测 API 云平台。云平台服务于新东方、流利说、沪江网等英语教育头部客户，提供面向终端用户的全方位英语口语评测功能。

项目 / 研究经历

问向学生关爱平台 AI 教师助手 *2025.02 –2025.06*

AI 教师助手通过分析学生心理数据，预警异常情况，自动筛选高风险学生并生成个性化关爱辅导策略方案，助力教师精准干预。在学生关爱案例的检索与生成场景中，检索准确率（Hits@3）提升至 0.93。

技术栈：*Python/Dify/LLM/Agent/MCP/TableStore/Milvus/LangGraph/Ollama/K8s*

- 搭建 RAG 知识库，支持学生画像分析与案例检索，生成个性化辅导方案。
- 设计混合检索架构，主导检索流程与核心模块开发，包括 TableStore/Milvus 技术选型。
- 开发语义匹配模块，基于 ICL（In-Context Learning）策略优化 Milvus 向量检索效果。
- 实现表格 Schema 解析，支持查询意图识别、检索模式判断及关键信息抽取。
- 设计 MCP Server 框架，完成与 Dify 平台的服务对接。
- 开发 Python 程序实现将 Dify 的工作流定义 DSL 文件转 LangGraph 代码工具。
- 负责使用 Ollama 搭建大模型及 Embedding 模型服务，并构建 Docker 镜像在 K8s 平台上部署和管理。

学生画像的宽表增量优化 *2024.10 –2024.11*

学生画像数据采用宽表结构（1000+ 列多维数据）支持个性化分析与建模，原全量日批处理方案资源消耗高且耗时，改进后转为日增量计算，冗余计算大幅减少，DataWorks 日均成本由 300 元降至 2 元。

技术栈：*DataWorks/Python/PySpark/SQL*

- 设计宽表转窄表的行转列逻辑，按画像维度及量表类型拆分结构化数据。
- 构建字段级变更识别机制，实现每日数据差异提取与增量计算。
- 使用 Python + SQL 实现增量数据与历史画像拼接流程，确保数据一致性与完整性。
- 优化调度链路，改造全量跑批为增量的调度逻辑。

基于 MLflow 的 AI 教练评分结果一致性与稳定性评价框架 *2024.09 –2024.10*

该项目开发了基于 MLflow 的自动化评分评估框架，用于分析 AI 教练在不同 Prompt 版本下的评分一致性和输出稳定性，将迭代验证周期从 3 天缩短至 2 小时，显著提升测试效率和迭代速度。

技术栈：*Python/MLflow*

- 设计并实现基于 MLflow 的评分输出日志自动化采集与管理流程。
- 构建评分一致性与稳定性评估指标体系，实现自动化对比分析报告生成。
- 优化 Prompt 迭代验证流程，实现版本迭代结果可追溯、可复现。

基于预训练语言模型的企业财务报告问答系统（研究项目） *2023.12 –2025.02*

该企业财报智能问答系统解析财报中的文本-表格混合数据并结合财务知识进行数值推理，提升分析效率和准确性。

技术栈：*Python/Pytorch/vLLM/deepspeed/llama-factory/llama-index/LoRA*

- 负责问答系统的界面、模块流程与交互功能设计与实现。
- 使用 deepspeed、llama-factory 通过 MiniCPM 2B、Qwen 2.5-1.5B 在 Multihiertt 数据集上 LoRA 微调（Supervised Fine-Tuning）。
- 提出 FGKI-CoT 和 EOSC 方法，使得参数量仅 1.5B 的 LLM 在 Multihiertt 数据集上数值推理的 EM/F1 达 55%/56%，CodaLab 打榜排名第一。
- 基于 llama-index 与 RAG 向量化处理混合数据，微调 BGE Embedding 模型将证据检索的 Hits@30 指标提升至 0.9 以上。
- 负责基于 vLLM 优化大模型推理服务，显著提升推理速度与吞吐量。
- 提交一项专利申请，两项软著，一篇 CCF C 类论文。

基于 BERT 的企业名称对齐方案设计与实施 *2023.04 –2023.08*

本方案采用基于 *BERT* 的分布式匹配技术，解决银行信用卡数据库中不规范企业名称（缩写/简写）与工商标准名称的对齐问题，将千万级数据的匹配准确率从 *60%* 提升至 *90%* 以上，计算效率提高 *3* 倍。

技术栈：*Python/PySpark/SimBERT*

- 负责需求分析工作、审查并理解兴业银行信用卡中心的企业名称数据特点。
- 设计与实施层级聚类方式以降低计算量。
- 设计并实现基于 Spark 并结合 BERT 的分布式计算程序，用于处理大规模数据的相似度计算。
- 负责性能优化工作：1. 优化确定降维参数　　2. 优化并确定相似度决策边界。

同盾云图产品架构和性能优化 *2021.08 –2023.03*

云图是基于知识图谱和大数据技术的金融级平台产品，通过构建业务场景知识图谱实现智能决策，广泛应用于反欺诈、风险管理等领域，提升商业决策效率。

技术栈：*Python/Spark Streaming/Scala/SQL/JanusGraph/TinkerPop/K8s*

- 解决平安项目中 Spark Streaming 的处理效率问题，将数据解析过程的处理效率提升数十倍。
- 梳理图谱中与 N 度关联相关的指标开发方法、计算逻辑和工作流程，重新制定技术实现方案，主导 Python GraphFrame 开发图指标计算工程代码。
- 主导设计平安项目中股权穿透图增量计算方案，实现增量数据处理策略。
- 主导设计使用 TinkerPop 作为内存图数据库实现图指标计算方案，核心代码开发和功能实现。
- 主导设计基于增量数据的图挖掘算法流程，设计和开发算法，应用于基于增量数据的连通图计算，优化股权穿透算法的执行效率提升 42 倍。
- 将基于增量数据的图挖掘算法应用于所有以连通图计算为前提的其他算法，如环形检测、金字塔识别算法等，优化算法执行效率提升数十倍。
- 主导兴业银行知识图谱项目相关服务的云原生（K8s）改造工作。

精锐教育集团数据门户 *2020.04 –2021.07*

针对精锐 *UPC* 系统数据报表分散、使用效率低的问题，主导重构数据门户，通过业务指标整合、权限精细化管控及自助分析功能升级，提升数据服务能力。

技术栈：*Java/Saiku/Mondrian/SQL*

- 负责重构报表体系，按业务域重组指标，优化表头及维度设计，提升使用效率。
- 设计行级和字段级权限管控体系，实现敏感数据的分级访问控制。
- 基于 Saiku+Mondrian 二次开发，集成 Kylin 实现拖拽式多维分析功能。

- 封装核心代码形成 SDK，完成 MDX 表达式转换引擎开发。
- 开发停课激活、结课等复杂业务指标计算逻辑。

基于 IRT 的学生能力值评估实时计算和模型中台 *2019.04 –2020.04*

针对智适应推荐引擎与业务系统高度耦合导致的迭代效率低、资源复用率差等问题，主导构建算法能力中台，通过微服务架构实现核心能力的标准化输出，支撑个性化推荐、实时学情分析等业务场景。

技术栈：*Java/Spring Boot/Spark Streaming/Hbase/Docker/K8s*

- 基于 Spring Boot 实现算法能力的服务化，完成用户画像服务（知识点掌握率、攻克率和学习效率标签）、实时推荐 API 等核心模块构建。
- 主导 Docker 容器化改造及 K8s 集群部署方案，实现服务自动扩缩容。
- 设计 Spark Streaming 实时处理框架，支撑千万级行为数据的即时分析与中间结果存储。
- 开发学生行为追踪系统，建立端到端数据链路监控能力。

松鼠 AI 学习系统图谱平台 *2018.04 –2019.04*

该平台将知识图谱从 *MySQL* 迁移至 *JanusGraph* 以提升图数据处理性能，并通过图谱服务平台 *API* 支撑推荐引擎、建课和 *CMS* 系统的调用。

技术栈：*Java/Spring Boot/JanusGraph*

- 主导图谱平台的相关技术调研选型、方案设计、API 接口设计和开发。
- 负责 JanusGraph 图数据库存储的研发工作。
- 负责引擎针对图操作的业务需求整理，包括搜索子图、知识点前后置、图谱环状检查，以及搜索 N 度关系等。
- 负责 Gremlin 语句的编写，将图操作的业务用 Gremlin 语句实现相关功能。

予见汽车金融数据和模型平台 *2017.02 –2018.04*

该平台整合多维度数据，通过可视化监控、智能风控体系和全流程管理，实现审批效率提升 *40%* 和风险识别准确率提升 *30%*。

技术栈：*Hadoop/Spark Streaming*

- 搭建 Hadoop+Spark Streaming 实时处理架构，实现用户行为数据的采集、计算和存储。
- 设计开发数据服务层，提供 BI 报表 (进件量/审批率等) 和用户行为分析 API。
- 构建关联人网络分析体系，通过 Neo4j 实现联系人关系图谱分析。
- 开发客户画像系统，设计 10+ 风险标签 (黑中介识别、敏感词检测、团伙关系标签等)。

予见汽车金融平台工作流引擎 *2017.08 –2018.04*

该平台通过可配置化工作流引擎实现信贷审批流程灵活调整，将风控规则迭代周期从 *1* 周缩短至 *1* 天，显著提升审批效率和业务响应速度。

技术栈：*Java/Spring Boot/Activiti*

- 重构 Activiti 源码适配金融业务需求，扩展核心功能。
- 抽象消息服务为可配置模块，支持流程节点动态嵌入。

“车萝卜”智能语音车载 HUD 产品 *2014.05 –2017.02*

“车萝卜”是国内第一款智能语音车载 *HUD* 导航终端，能够解放司机的双手，实现语音操控完成导航，打电话等功能。其中 *HUD* 的功能采用翻转屏幕镜像投影反射的方式，将导航、路况、车辆信息呈现在前挡风玻璃中，在不影响司机正常驾驶的情况下进行交互，提高驾驶安全性。

技术栈：*C/C++/Android HAL/驱动开发*

- 主导“车萝卜”产品的 DLP 投影光机，OBD 串口接口，FMTX 无线投射模块，胎压传感器的驱动程序开发工作。

- 负责光感传感器配合光机亮度调节的相关测试工作，参数验证工作。
- 负责编写 Android HAL 层模块，编写 native 服务，实现调用设备驱动文件接口，实现 Binder 与 Android App 通信机制。
- 负责“车萝卜”二代青春版 OBD 功能实现，完成固件升级服务。

语音识别与评测后端 API 平台 *2011.07 –2014.04*

负责基于 *Red5* 流媒体服务器和 *Spring* 框架开发工作，提供音频流处理服务。平台接收前端 *Flash* 通过 *RTMP* 传输的音频流，将其拆包并封装为 *HTTP* 请求，异步发送至后端计算节点。计算节点由多个语音识别/评测算法服务组成，通过 *Proxy* 服务实现负载均衡。识别结果返回客户端并存储至 *MongoDB*。

技术栈：*Java/Spring/MongoDB*