

张威健 | TIM ZHANG

数据工程师 | AI 驱动的数据管线架构师 | Data + AI 系统构建者
weijianlucky@gmail.com | HK: +852 84965467 | Mainland: +86 131 6809 0613
Portfolio: portfolio.timzhang.cn | GitHub: github.com/February13

职业概要

8 年以上生产级数据工程经验，覆盖商业银行、跨国零售、云数仓、公共服务自动化、保险反欺诈与临床 AI。曾交付 10TB+ 数仓上云、五层数仓建模、SOX 友好的财务数据流程、StarRocks 查询加速与生产报表管线。近年工作进一步延伸到 Agentic RAG、GraphDB 与临床 AI：代表项目包括获岭南杯二等奖的 SRR 智能体案件处理系统、GaitGPT 临床步态分析智能体，以及 MCC FWA 保险反欺诈图谱智能系统。

科研成果、专利与奖项

论文准备中：From Symptoms to Signals: A Bidirectional LLM Framework for Literature-Grounded Clinical Gait Analysis.

专利交底准备中：File-Driven Intelligent Case Processing System and Method for Public-Service Complaint Workflows / 一种面向公共服务投诉流程的文件驱动智能案件处理系统及方法。

发明披露准备中：Evidence-Bound Bidirectional Clinical Gait Translation System and Method.

专利交底准备中：视联发现：跨模态视觉关系的多模型融合方法。

岭南杯二等奖。第四届中山大学“岭南杯”中国经济发展案例分析大赛；香港岭南大学 MScAIBA 团队 SRR 案例从 163 支队伍中入围 8 支决赛队伍。

教育经历与 AIBA 课程

人工智能与商业分析硕士 | 香港岭南大学 | 2025 - 2026

- AIBA 相关课程：人工智能基础、商业数据管理、数据分析与编程原理、区块链、医疗健康分析、数据可视化、生成式 AI 编程、AI-Based Optimization。

管理学学士 | 深圳大学 | 2014 - 2017

技术能力

编程与数据：Python, SQL, SparkSQL, Impala SQL, Bash/Shell, Java

数据工程：Hadoop, Spark, Hive, StarRocks, Kafka, Seatunnel, DolphinScheduler, Informatica PowerCenter

AI/LLM 系统：FastAPI, LangChain/LangGraph, RAG, Agentic workflows, pgvector, GraphDB/Neo4j, OpenAI GPT-4o, Gemini, LLM-as-Judge

金融与零售数据：银行数据模型、B2B 订单/合同/损益报表、理赔审核、SOX 友好的数据流程

云与 DevOps：Google Cloud Run, Cloud SQL, Tencent Cloud, Docker, Linux, GitHub Actions

代表性 AI 与数据项目

SRR 智能体案件处理系统 | 技术负责人 | 99.4% 贡献度 | 岭南杯二等奖 | 电视报道 | 2025-2026

项目链接：[主页案例](#) | [公开仓库](#)

- 项目背景：香港 SRR 案件处理依赖多渠道材料、人工分派、历史记录查找与 A-Q 字段反复解释，流程长且可追溯性要求高。
- 责任描述：设计 7 层智能体架构与 17 个原子能力，覆盖 ICC 1823/TMO/RCC 解析、A-Q 抽取、部门路由、pgvector + RRF 检索、回复草拟、评估与回滚。
- 项目成果：完成 12/12 项需求与 80,430 行内部迭代代码；沉淀为获奖案例、电视报道、公开参考实现与专利交底准备材料。

技术栈：Python, FastAPI, React/TypeScript, PostgreSQL 15 + pgvector, GPT-4o, Google Cloud Run, Cloud SQL, Docker

GaitGPT：文献增强临床步态分析智能体 | 架构负责人 | 论文准备中 | iCAN 2026 | 岭南大学发明披露 | 2026

项目链接：[主页案例](#) | [公开仓库](#)

- 项目背景：步态数据提供密集传感器指标，而临床问题通常以跛行、拖步、不对称等症语言提出，需要在症状与可测量指标之间建立桥梁。
- 责任描述：负责系统方向与架构设计，完成反向翻译、知识解释、PubMed/Semantic Scholar 检索、证据评分、规则模板优先查询与物理约束校验。
- 项目成果：在 20 题 NONSD-Gait benchmark 上达到 0.9223 综合准确率与 10.07 秒中位响应时间；准备论文 From Symptoms to Signals、iCAN 2026 与发明披露材料。

技术栈：Python, FastAPI, Next.js/React, SurrealDB, LangGraph, PubMed/Semantic Scholar APIs, Ollama/Qwen, OpenAI-compatible providers

MCC FWA：保险理赔反欺诈图谱智能系统 | 核心开发 | 企业协作 | InsurTech / Healthcare AI | 2026

- 项目背景：保险 FWA 审核中欺诈标签稀疏，但患者、机构、诊断、票据、计费项目和金额模式之间存在大量关系型风险信号。
- 责任描述：将 38,659 份医疗理赔 JSON 转换为属性图谱，形成 161 万节点与 257 万边；建设可迁移 nodes/edges CSV 抽取、Neo4j 路径校验与 FastAPI 异步审核能力。
- 项目成果：将审核界面从黑箱 approve/reject 转为可审计的证据链、风险原因、相关实体、保单校验与 claim 级解释。

技术栈：Python, FastAPI, JSON/CSV pipelines, Neo4j, Google Spanner Graph-ready design, LangChain/LangGraph, OpenRouter LLMs

企业级数据工程经历

数据工程师 | 深圳市仁锐人力科技有限公司 | 09/2023 - 09/2025

- 项目背景：跨国零售企业的 B2B 财务与运营报表依赖订单、合同、账单、库存、礼品卡、仓储和财务等碎片化数据源。
- 责任描述：迁移 10TB+ IDC 数仓至腾讯云，重建 ODS/DIM/DWD/DWS/ADS 分层；交付项目、订单、商品、客户、门店、履约、库存、毛利、支付、审批与 ETL 装载字段的 ADS 报表表，并配套 SOX 友好发布证据与生产监控。
- 项目成果：Seatunnel 同步一致性达到 99.95%，故障减少 70%；ORC/SparkSQL 优化使磁盘占用降低 38%、查询速度提升 55%、ETL 时长降低 87.5%；StarRocks + Hive 外表加速将核心分析查询从 3-10 分钟降至平均 0.8 秒。

数据工程师 | 杭州亚信拓科技有限公司 | 04/2021 - 08/2023

- 项目背景：企业报表团队需要在 100+ 交易系统上沉淀可复用的数仓模型层，并统一指标口径。
- 责任描述：设计 5 层 Hive 数仓模型，使用 Informatica PowerCenter 集成源数据，开发 Impala/HQL 指标逻辑与 UDF/UDTF 处理。
- 项目成果：提升跨团队模型复用约 50%，并通过敏感数据脱敏与复用计算逻辑支持合规分析。

数据工程师 | 深圳市赞同科技股份有限公司 (客户：商业银行) | 02/2019 - 03/2021

- 项目背景：商业银行数据集市需要稳定定义公司信贷、额度、借款合同、放款、还款、结算账户和产品级报表口径。
- 责任描述：维护数仓模型层 SQL，将集市需求转化为可追溯逻辑，并与银行业务方对齐源表血缘、问题处理与口径变更。
- 项目成果：支撑财务、监管与管理报表输出，形成一致定义、受控变更和稳定下游交付。

工程师 | 成都华讯中星科技有限公司 | 07/2017 - 11/2018

- 项目背景：IBM 应用运维需要快速故障响应与可重复支持流程。
- 责任描述/成果：处理网络与系统问题，并编写 Bash/Python 自动化脚本减少重复人工操作。

语言能力

语言：粤语 (母语)、普通话 (母语)、英语 (职业工作能力)