TIM ZHANG

Data Engineer | AI-Driven Pipeline Architect | Data + AI Systems Builder
weijianlucky@gmail.com | HK: +852 84965467 | Mainland: +86 131 6809 0613
Portfolio: portfolio.timzhang.cn | GitHub: github.com/February13

PROFESSIONAL SUMMARY

Data Engineer and AI-driven pipeline architect with 8+ years of production data engineering experience across commercial banking, multinational retail, cloud data warehouses, GovTech automation, insurance FWA, and clinical AI. Delivered 10TB+ cloud migration, five-layer warehouse architecture, SOX-aware financial data workflows, StarRocks performance optimization, and production reporting pipelines before moving into Agentic RAG, GraphDB, and clinical AI systems. Recent flagship work includes the award-winning SRR Agentic Case Processing System, GaitGPT clinical gait analysis agent, and MCC FWA insurance fraud graph intelligence system.

PUBLICATIONS, PATENTS & AWARDS

Zhang, W. (manuscript in preparation). From Symptoms to Signals: A Bidirectional LLM Framework for Literature-Grounded Clinical Gait Analysis.
Zhang, W. (patent disclosure in preparation). 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.
Zhang, W. (invention disclosure in preparation). Evidence-Bound Bidirectional Clinical Gait Translation System and Method.
Zhang, W. (patent disclosure in preparation). Vision-Link Discovery: Multi-Model Fusion Method for Cross-Modal Visual Relationship Discovery.
Lingnan Cup Second Prize. 4th Sun Yat-sen University Lingnan Cup China Economic Development Case Competition; HK Lingnan University MScAIBA team SRR case selected from 163 teams, with 8 finalists.

EDUCATION & AIBA COURSEWORK

MSc in Artificial Intelligence and Business Analytics | Lingnan University, Hong Kong | 2025 - 2026
- Relevant AIBA coursework: Foundation of Artificial Intelligence; Business Data Management; Principle of Data Analytics and Programming; Blockchain; Healthcare Analytics; Data Visualization; Programming with Generative AI; AI-Based Optimization.

Bachelor of Management | Shenzhen University, China | 2014 - 2017

TECHNICAL SKILLS

Programming & Data: Python, SQL, SparkSQL, Impala SQL, Bash/Shell, Java
Data Engineering: Hadoop, Spark, Hive, StarRocks, Kafka, Seatunnel, DolphinScheduler, Informatica PowerCenter
AI/LLM Systems: FastAPI, LangChain/LangGraph, RAG, Agentic workflows, pgvector, GraphDB/Neo4j, OpenAI GPT-4o, Gemini, LLM-as-Judge
Financial & Retail Data: Banking data models, B2B order/contract/P&L reporting, claims adjudication, SOX-aware data workflows
Cloud & DevOps: Google Cloud Run, Cloud SQL, Tencent Cloud, Docker, Linux, GitHub Actions

SELECTED AI & DATA PROJECTS

SRR Agentic Case Processing System | *Technical Lead | 99.4% contribution | Lingnan Cup 2nd Prize | Featured on TV | 2025-2026*
Project links: Portfolio case study | Public showcase repo
- Situation/Task: Hong Kong SRR case handling relied on multi-channel materials, manual routing, historical lookup, and repeated A-Q field interpretation.
- Action: Designed a 7-layer agentic architecture with 17 atomic capabilities covering ICC 1823/TMO/RCC parsing, A-Q extraction, department routing, pgvector + RRF retrieval, reply drafting, evaluation, and rollback.
- Result: Delivered 12/12 requirements and 80,430 lines across internal iterations; converted the system into an award-winning case with TV coverage, public reference code, and patent-prep narrative.

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

GaitGPT: Literature-Augmented Clinical Gait Analysis Agent | *Architecture Lead | Manuscript in preparation | iCAN 2026 | LU invention disclosure | 2026*
Project links: Portfolio case study | Public showcase repo
- Situation/Task: Gait datasets produce dense sensor indicators, while clinicians reason in symptom terms such as limping, shuffling gait, and asymmetry.
- Action: Led system direction and architecture; implemented reverse translation, knowledge explanation, PubMed/Semantic Scholar retrieval, evidence scoring, rule-template-first queries, and physics-constrained validation.
- Result: Evaluated on a 20-question NONSD-Gait benchmark with 0.9223 overall accuracy and 10.07s median response time; prepared manuscript "From Symptoms to Signals" plus iCAN 2026 and LU invention-disclosure materials.

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

MCC FWA: Insurance Claims Fraud Graph Intelligence System | *Core Developer | Enterprise collaboration | InsurTech / Healthcare AI | 2026*
- Situation/Task: Insurance FWA review has sparse fraud labels and relationship-heavy risk signals across patients, providers, diagnoses, receipts, billing items, and amount patterns.
- Action: Converted 38,659 medical claim JSON files into a 1.61M-node / 2.57M-edge property graph, modeled major claim entities, and built portable CSV extraction, Neo4j path validation, and FastAPI async review support.
- Result: Shifted the review surface from opaque approve/reject labels to auditable evidence trails, risk reasons, related entities, policy checks, and claim-level explanations.

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

ENTERPRISE DATA ENGINEERING EXPERIENCE

Data Engineer | Shenzhen RenruiHR Technology Co., Ltd. | 09/2023 - 09/2025

- Situation/Task: B2B finance and operations reporting for a multinational retail enterprise depended on fragmented order, contract, billing, inventory, gift-card, warehouse, and finance sources.
- Action: Migrated 10TB+ IDC warehouse workloads to Tencent Cloud, rebuilt ODS/DIM/DWD/DWS/ADS layers, and delivered report-ready ADS tables across project, order, item, customer, store, fulfillment, inventory, margin, payment, approval, and ETL-load fields.
- Result: Seatunnel sync reached 99.95% consistency and cut failures by 70%; ORC/SparkSQL optimization reduced disk by 38%, improved query speed by 55%, cut ETL runtime by 87.5%, and StarRocks + Hive external table acceleration moved core analytical queries from 3-10 minutes to avg. 0.8s.

Data Engineer | Hangzhou Yatop Information Technology Co., Ltd. | 04/2021 - 08/2023

- Situation/Task: Enterprise reporting teams needed reusable warehouse model layers across 100+ transaction systems and inconsistent metric logic.
- Action: Architected 5-layer Hive warehouse models, integrated source data with Informatica PowerCenter, and built Impala/HQL metric logic plus UDF/UDTF processing.
- Result: Improved cross-team model reuse by 50% and supported privacy-compliant analytics through sensitive-data masking and reusable computation logic.

Data Engineer | Shenzhen Owned Technology Co., Ltd. (Client: Commercial Bank) | 02/2019 - 03/2021

- Situation/Task: Commercial-bank marts needed stable definitions for corporate credit, limits, loan contracts, disbursement, repayment, settlement accounts, and product-level reporting.
- Action: Maintained warehouse model-layer SQL, translated mart requirements into traceable logic, and aligned source lineage / issue handling with banking stakeholders.
- Result: Supported financial, regulatory, and management reporting with bank-grade delivery discipline: consistent definitions, controlled changes, and stable downstream outputs.

Engineer | Chengdu HQtimes Technology Co., Ltd. | 07/2017 - 11/2018

- Situation/Task: IBM application operations required fast issue response and repeatable support routines.
- Action/Result: Troubleshot network/system issues and built Bash/Python automation scripts to reduce repetitive manual operations.

LANGUAGES

Languages: Cantonese (Native), Mandarin (Native), English (Professional Working Proficiency)