

Kaiwen Lin

Address: Austin, TX78705

Expected graduate time : May 2026

+1-7373267625

✉ Kaiwenlin@utexas.edu

🐙 GitHub Profile

EXPERIENCE

- The Institute of World Economics and Politics, Chinese Academy of Social Sciences** 2020–2023
Research Assistant Beijing
 - Built ETL pipelines (PySpark, PostgreSQL) integrating 20M+ UN trade records with real-time financial data via REST APIs (Wind), reducing processing time by 70% and powering cross-dataset policy dashboards.
 - Optimized database schema and query performance, enabling concurrent access for 10+ analysts and sub-second response times for complex trade flow queries.
 - Co-authored 2 CSSCI publications on global trade patterns as part of a National Social Science Foundation project; received "Top 10 Paper of the Year" and 2 other research awards.

PROJECT

- Distributed Message Queue in Go** Nov 2024 – March 2025
 - Built a message broker in Go achieving 8K msgs/sec throughput with topic-based routing and at-least-once delivery via write-ahead log and consumer acknowledgments; deployed 3-node cluster with leader election for high availability.
 - Designed TCP binary protocol with flow control (RDY mechanism) to prevent consumer overload; implemented message replication across followers with configurable durability levels (async/sync replication).
 - Optimized throughput from 2.5K to 8K msgs/sec by reducing mutex contention with channel-based concurrency and batching disk writes (50ms windows); exposed Prometheus metrics for message lag and node health monitoring.
- Car Accident & Insurance Claims Data Lakehouse on GCP** Aug 2025 – Present
 - Architected a modern data lakehouse on GCP, ingesting structured and unstructured sources (PDF, text reports) into Cloud Storage; leveraged LLMs (Gemini API) to extract structured attributes from unstructured accident narratives, enriching analysis with NLP-derived features.
 - Built production-grade dbt pipelines orchestrating incremental transformations across staging, intermediate, and mart layers in BigQuery; implemented dimensional modeling (star schema) and data quality tests (dbt-expectations).
 - Unified disparate datasets through semantic modeling, enabling cross-source analytics; delivered BI-ready marts supporting complex queries with sub-second performance via partitioning and clustering optimization.

EDUCATION

- The University of Texas at Austin** 2024 - Expected May 2026
Master of Information Science GPA: 4.0/4.0;
 - Courses:** Algorithm; Deep Learning and Multimodal Systems; JavaScript; Distributed System
- University of Chinese Academy of Social Sciences** 2020 - 2023
Master of International political economy Graduate with Distinction, GPA: 3.9/4.0(Overall); 4.0/4.0(Core)
- Dalian University of Foreign Languages** 2016-2020
Bachelor of International Affairs and International Relations Graduate with Distinction

HONORS&AWARDS

- Top 10 Best Chinese Papers in Global Governance in 2022, The Yearbook of World Economy 2023
- China National Scholarship (0.2%), Ministry of Education of China 2022
- Outstanding Graduate(10%),Comprehensive First/Second Class Scholarship, UCASS 2021-2023

SKILLS

Programming Languages: Java, Golang, Python, JavaScript, R, SQL, C++

Frameworks: PySpark, React, Node.js, Spring Boot, Airflow, PyTorch, OpenCV

DevOps & Cloud: Google Cloud Platform, Git, CI/CD, Postman, Tableau

Techniques: Fine-tuning, NLP, Time-Series Analysis, Bayesian Modeling, Causal Inference

Languages:Mandarin Chinese(native); English (fluent)