




Kaiwen Lin Address: Austin, TX78705 Expected graduate time : May 2026		 +1-7373267625  Kaiwenlin@utexas.edu  GitHub Profile
EDUCATION		
<ul style="list-style-type: none"> The University of Texas at Austin <i>Master of Information Science</i> –Algorithm; Deep Learning and Multimodal Systems; Data Warehouse; Bayesian statistical methods University of Chinese Academy of Social Sciences <i>Master of International political economy</i> Graduate with Distinction, GPA: 3.9/4.0(Overall); 4.0/4.0(Core) Dalian University of Foreign Languages <i>Bachelor of International Affairs and International Relations</i> Graduate with Distinction 	2024 - Expected May 2026 GPA: 3.9/4.0; 2020 - 2023 2016-2020	
PROJECT		
<ul style="list-style-type: none"> Enterprise Financial Data Integration & Quality Platform <i>AWS (MWAA, Glue, Redshift, S3), Scala, Spark, Amazon Deequ, KMS, Terraform, Slack</i> – Orchestrated a Medallion Architecture data lakehouse on AWS using MWAA (Airflow) to process daily equity records via Scala Spark jobs, ensuring strict idempotency and data lineage across Bronze, Silver, and Gold layers. – Developed robust ETL pipelines in AWS Glue with Amazon Deequ as a declarative quality gate; implemented a circuit-breaker mechanism to automatically abort job execution upon exceeding a 5% threshold of constraint violations or schema anomalies, preventing downstream data pollution. – Optimized large-scale storage performance by implementing date-based partitioning in Parquet format on S3, leveraging KMS-managed encryption and IAM service-linked roles to maintain strict multi-tenant resource isolation and security compliance. – Engineered an automated warehouse sink into Redshift Serverless, utilizing S3 lifecycle policies for cost-efficient data aging and configuring real-time failure notifications via CloudWatch and Slack for proactive pipeline monitoring and recovery. Car Accident & Insurance Claims Data Lakehouse on GCP <i>Google Cloud Platform(BigQuery, IAM), dbt, Python, LangChain, Data Lakehouse, Dimensional Modeling</i> – Architected a 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). 	Nov 2025 – Present Aug 2025 – Oct 2025	
EXPERIENCE		
<ul style="list-style-type: none"> The Institute of World Economics and Politics, Chinese Academy of Social Sciences <i>Research Assistant (Data Platform / Data Science)</i> –Architected an ETL pipeline using Apache Airflow to orchestrate the ingestion of 500GB+ heterogeneous global trade and financial datasets (UN Comtrade, IMF, Wind) into AWS S3; implemented monitoring and daily API ingestion workflows that reduced data preparation lead time by 85%. –Led the migration of legacy research data to Snowflake Cloud Data Warehouse, designing optimized relational schemas and partitioning strategies that achieved a 90% reduction in query latency for complex, multi-billion-row econometric joins. –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. 	2020–2023 Beijing	
HONORS&AWARDS		
<ul style="list-style-type: none"> Top 10 Best Chinese Papers in Global Governance in 2022, The Yearbook of World Economy China National Scholarship (0.2%), Ministry of Education of China Outstanding Graduate(10%),Comprehensive First Class Scholarship, UCASS 	2023 2022 2021-2023	
SKILLS		
Programming Languages: Java, Python, Scala, SQL, Golang, R Data Engineering: Spark (Scala/Python), Airflow, dbt, Amazon Deequ, Pandas, PyTorch Cloud & Data Warehouse: AWS (S3, Glue, Redshift, MWAA), GCP (BigQuery, GCS), Snowflake, Terraform DevOps & Tools: Git, Docker, CI/CD, Splunk, Tableau, Swagger Techniques: Fine-tuning, NLP, Time-Series Analysis, Bayesian Modeling Languages: Mandarin Chinese(native); English (fluent)		