

MIN (MIA) SHI

AI Engineer and Applied Data Scientist with a Ph.D. background, specializing in AI-driven data pipelines, LLM-based applications, and analytics systems supporting public health and policy decision-making. Experienced in deploying production AI systems on AWS and applying research insights to real-world governance challenges.

minshimia@gmail.com ◇ Min.Shi@utdallas.edu ◇ +1 (469) 403-7557
Google Scholar ◇ ResearchGate ◇ LinkedIn ◇ Personal Portfolio

Education

The University of Texas at Dallas Ph.D. in Political Science – Quantitative Statistical Modeling Focused	Aug. 2019 – Dec. 2024 GPA: 3.95/4.0
The University of Texas at Dallas M.S. in Social Data Analytics and Research	Aug. 2021 – Aug. 2024 GPA: 3.95/4.0
The University of Texas at Dallas M.S. in Business Analytics (Data Science & Data Engineering Track)	Aug. 2022 – May 2024 GPA: 4.0/4.0
The University of Texas at Dallas Graduate Certificate in Applied Machine Learning	Aug. 2022 – May 2023 GPA: 4.0/4.0
The University of Texas at Dallas M.A. in Political Science	Aug. 2019 – May 2022 GPA: 3.95/4.0
Shandong University M.L. in International Politics	Sept. 2016 – Jun. 2019 GPA: 88.78/100
Daito Bunka University Exchange Student in Political Science	Sept. 2017 – Aug. 2018
Shandong University B.A. in Japanese	Sept. 2012 – Jun. 2016 GPA: 87.37/100

Scholarships

Keith Lankford Tayer Fellowship	2024
John Forrest Kain Scholarship	2023
Government and Political Science Scholarship	2022

Work Experience

Reframe Data Services <i>AI Engineer & Data Scientist</i> → North Bethesda, MD	May 2025 – Present
• Owned quality assurance and evaluation of production GenAI agents integrating multiple LLM providers (Claude, OpenAI, Gemini), ensuring reliability, traceability, and alignment with organizational standards.	
• Designed automated evaluation frameworks for LLM outputs, including hallucination detection, rule-based validation, accuracy benchmarking, and cost monitoring across 17+ OpenSearch-backed RAG pipelines.	
• Applied governance controls to AI pipelines (prompt versioning, output constraints, audit logging), supporting safe deployment of decision-facing AI systems.	
• Built CI/CD-enabled GenAI pipelines on AWS (ECS, EKS, ECR, Lambda, S3) with automated testing and rollback, achieving 99.9% uptime while reducing inference cost by 68% and latency by 44%.	
• Collaborated cross-functionally with product leaders, editors, and engineers to validate AI behavior against business and policy requirements, improving analyst productivity by 40%.	

The Sunwater Institute <i>Data Engineer</i> → North Bethesda, MD / Remote	Jun. 2024 – May 2025
• Developed and validated large-scale NLP and speech-processing pipelines (AWS Transcribe, Textract) achieving 90%+ transcript accuracy for policy-critical datasets.	

- Implemented automated data quality checks, schema validation, anomaly detection, and failure monitoring to support downstream ML evaluation and reporting accuracy, reducing pipeline errors by 75%.
- Built ETL pipelines (Python, PySpark, SQL) with reproducibility and auditability to support downstream ML and analytics use cases.
- Partnered with researchers and leadership to ensure analytical outputs met documentation, quality, and governance expectations.

The University of Texas at Dallas Data Analyst & Research Assistant

May 2020 – May 2024

→ Richardson, TX / Part-time

- Designed, evaluated, and validated 20+ predictive, NLP, and time-series models supporting applied social-policy and global-health analytics.
- Applied text mining, semantic similarity, and statistical methods to large-scale survey and unstructured text data, contributing to peer-reviewed publications.
- Translated complex analytical results into stakeholder-ready insights and visual summaries for interdisciplinary research teams.
- Led and mentored a team of five research assistants, overseeing data collection, quality checks, modeling workflows, and delivery timelines.

Lucion Technology Corp., Ltd. Marketing Data Analyst (Intern)

Jul. 2017 – Aug. 2017

→ Jinan, China

- Improved data extraction efficiency by 40% through MySQL optimization and automation.
- Produced BI reports and competitor insights to support marketing strategy and product understanding.

Publications

Peer-Reviewed Journal Articles

- Palifka, B. J., Hanson-DeFusco, J., **Shi, M.** (2026). Perceptions of Corruption as a Factor in Academic Integrity among University Students in Mexico and the United States. doi:10.21203/rs.3.rs-7235522/v1
- Shah, H., Hanson-DeFusco, J., Popalzai, H., *et al.*, **Shi, M.** (2025). Diminished Quality of Life and Psychosocial Strain of Women Under the New Taliban Era: A Thematic Analysis. *Societies*, 16(1), 9. doi:10.3390/soc16010009
- Hanson-DeFusco, J., Sobolov, A., Stanekzai, S., *et al.*, **Shi, M.** (2025). The association of diminished quality of life of Afghan adults' psychosocial wellbeing in the era of the Taliban 2.0 government. *PLOS Mental Health*, 2(1), e0000118. doi:10.1371/journal.pmen.0000118
- Hanson-DeFusco, J., **Shi, M.**, Du, Z., *et al.* (2023). Systems analysis of the effects of the 2014–16 Ebola crisis on WHO-reporting nations' policy adaptations and 2020–21 COVID-19 response: a systematized review. *Globalization and Health*, 19, 96. doi:10.1186/s12992-023-00997-8

Other Publications (Policy/Regional)

- Yang, L., **Shi, M.** (2023). An Analysis of the Motivation of the Abe Cabinet's Policy Evolution and Adjustment towards China. *Journal of China's Neighboring Diplomacy*, 6(2), 43–68. Available via ResearchGate
- Yang, L., **Shi, M.** (2019). China Policy Adjustment or Changes by the Abe Administrations and Its Impacts. *Peace and Development*, (3), 66–84. Available via ResearchGate

Research Experience

The University of Texas at Dallas Research Assistant

May 2020 – May 2024

→ Multiple Faculty

- Supported 10+ global health/policy projects using advanced statistical modeling (fixed-effects regression, difference-in-difference, time series) and NLP.
- Managed data collection via Qualtrics surveys and web scraping using Python and R; maintained reproducible analysis workflows.
- Led and coordinated a team of five junior assistants, improving collaboration and on-time delivery across concurrent projects.

Conferences & Presentations

ISDSA Annual Meeting	<i>Presenter</i>	Jul. 2025
→ Washington, DC		
The Waves of US-China Technology Decoupling and Its Implication on MNCs: Using VAR Time Series Approach		
MPSA Annual Conference	<i>Presenter</i>	Apr. 2024
→ Chicago, IL		
The Waves of US-China Technology Decoupling and Its Implication on MNCs: Using VAR Time Series Approach		
SPSA Annual Conference	<i>Presenter</i>	Jan. 2024
→ New Orleans, LA		
Navigating the Waves of US-China Decoupling: A Comparative Analysis of Trade and Investment in Technology and Non-Technology Sectors		
ISDSA Annual Meeting	<i>Presenter</i>	Jul. 2023
→ Shanghai, China		
China's COVID Lockdown Policy and Trade with US: A Deep Learning Time Series Approach		
ISDSA Meeting	<i>Presenter</i>	May–Jun. 2022
→ Notre Dame, IN		
Modeling US-China Trade Relations: A Time Series Machine Learning Approach Using MNC Stock Data		

Teaching Experience

The University of Texas at Dallas	<i>Graduate Teaching Assistant</i>	Aug. 2019 – May 2024
→ EPPS / Political Science		
Supported undergraduate and graduate courses in American government, public policy, international relations, quantitative methods, and cybersecurity policy. Delivered lectures when needed, facilitated discussion sections, and supported assessment development and grading.		

Selected AI & Data Projects

AI-Powered Payment Service Chatbot	Aug. 2023 – Dec. 2023
• Built an NLP-driven chatbot (Python, MySQL) over 10M+ records; improved response efficiency by 25% and achieved 99% model accuracy (XGBoost).	
Kaggle Plant Pathology (Deep Learning)	Nov. 2023 – Dec. 2023
• Fine-tuned ConvNeXt with transfer learning; achieved 86.8% accuracy and Top-3 ranking.	

Technical Skills

AI & Machine Learning	LLMs; RAG; NLP; Conversational AI; Deep Learning; Machine Learning; MLOps; PyTorch; TensorFlow
Data Engineering & Analytics	ETL/ELT; Data Orchestration; Apache Spark; GraphQL; OpenSearch; PostgreSQL; MySQL; SQL Server; AWS RDS
Cloud & Infrastructure	AWS (ECS, Lambda, S3, Glue, EC2, OpenSearch, Parameter Store, CloudWatch); Docker; Kubernetes; CI/CD; Jenkins; GitHub Actions; Microservices
Programming Development	& Python; SQL; Node.js; FastAPI; API Development; React; Streamlit; TypeScript; Java; R; Stata; SAS
Visualization Apps	& Dash; Tableau; Power BI; R Shiny
Automation Low-Code Certifications	& Alteryx; Appian; ACCELQ; UiPath
Languages	Graduate Certificate in Applied Machine Learning (UT Dallas); Google Data Analytics; AWS Certified Cloud Practitioner; Alteryx Designer Core; Appian Certified Associate Developer; ACCELQ Automation Engineer English; Chinese; Japanese