MIN (MIA) SHI

Cecil H. Green Hall | 800 West Campbell Road, GR 3.314 | Richardson, TX 75080-3021
Min.Shi@utdallas.edu ◇ minshimia@gmail.com
LinkedIn ◇ Personal Portfolio

Education

The University of Texas at Dallas — **Aug. 2019 – Dec. 2024**
Ph.D. in Political Science – Quantitative Statistical Modeling Focused — *GPA: 3.95/4.0*

The University of Texas at Dallas — **Aug. 2021 – Aug. 2024**
M.S. in Social Data Analytics and Research — *GPA: 3.95/4.0*

The University of Texas at Dallas — **Aug. 2022 – May 2024**
M.S. in Business Analytics (Data Science & Data Engineer Track) — *GPA: 4.0/4.0*

The University of Texas at Dallas — **Aug. 2022 – May 2023**
Graduate Certificate in Applied Machine Learning — *GPA: 4.0/4.0*

The University of Texas at Dallas — **Aug. 2019 – May 2022**
M.A. in Political Science — *GPA: 3.95/4.0*

Shandong University — **Sept. 2016 – Jun. 2019**
M.L. in International Politics — *GPA: 88.78/100*

Daito Bunka University — **Sept. 2017 – Aug. 2018**
Exchange Student in Political Science

Shandong University — **Sept. 2012 – Jun. 2016**
B.A. in Japanese — *GPA: 87.37/100*

Scholarships

Keith Lankford Tayer Fellowship — **2024**

John Forrest Kain Scholarship — **2023**

Government and Political Science Scholarship — **2022**

Teaching Experience

School of Economic, Political and Policy Sciences, UTD *Teaching Assistant* — **Aug. 2019 – May 2024**

Graduate Teaching Assistant for:
GOVT 2305 American National Government
GOVT 2306 State and Local Government
EPPS 2302 Methods of Quantitative Analysis in the Social and Policy Sciences
PSCI 3301 Political Theory
PSCI 3325 American Public Policy
PSCI 3328 International Relations
PSCI 3350 Comparative Politics
PSCI 4348 Terrorism, PPOL 4396 Topics in Public Policy
PPPE 6302 Conflict in Cyberspace
PPPE 6304 Open-Source Intelligence for Cyber Security and Policy
EPPS 6313 Introduction to Quantitative Methods

Responsibilities as followed:
- Assisted in course preparation and delivered lectures for multiple courses across political science and policy studies.
- Managed and conducted seminars to disseminate research findings and facilitate knowledge sharing among students.
- Stepped in to cover lectures for professors during their absence, ensuring seamless course continuity.
- Played a key role in organizing and executing various academic events and workshops aimed at enhancing student engagement and learning in topics like public policy.
- Contributed to the development and grading of course assessments, providing timely and constructive feedback to students.
- Oversaw class attendance and participated in departmental meetings to contribute to curriculum development and

course planning.
• Provided individualized assistance to students, addressing queries and guiding them in Methods of Quantitative Analysis and Introduction to Quantitative Methods.
• Coordinated with faculty members to streamline course content and materials for complex subjects like Open-Source Intelligence for Cyber Security and Policy.
• Actively involved in academic mentoring, guiding students through complex policy and political concepts, and helping them apply these in real-world contexts.
• Assisted in research activities, enhancing the course content for Topics in Public Policy and Comparative Politics with updated, relevant information.

Research Experience

School of Economic, Political and Policy Sciences, UTD *Research Assistant* **May 2022 – May 2024**
↪ Prof. Jessi Hanson-Defusco

Took responsibility for data analysis for 10+ global health/policy projects using advanced statistical models.

• Managed data collection in diverse methods including Qualtrics surveys and web scraping using R and Python.
• Developed 20+ robust statistical models (multi-variable and fixed-effect regression, difference-in-difference, time-series) combined ML models and NLP skills to support correlation and causal inference in research.
• Led a team of five junior assistants, ensuring collaboration and timely project completion, underscoring strong communication, teamwork, leadership, and project management skills.

School of Economic, Political and Policy Sciences, UTD *Research Assistant* **May – Aug. 2021**
↪ Prof. Thomas Gray, Prof. Banks Miller

• Conducted data collection of 1291 Supreme Court cases using both manual and web-scripting techniques, ensuring accurate and comprehensive data capture.
• Utilized time-series models to analyze and assess the time gaps among the schedules of the court cases, providing insights into the temporal dynamics of the legal proceedings.

School of Economic, Political and Policy Sciences, UTD *Research Assistant* **May – Aug. 2020**
↪ Prof. Jonas Bunte

• Conducted data collection of 1291 Supreme Court cases using both manual and web-scripting techniques, ensuring accurate and comprehensive data capture.
• Utilized time-series models to analyze and assess the time gaps among the schedules of the court cases, providing insights into the temporal dynamics of the legal proceedings.

Conferences

2024 MPSA Annual Conference — Chicago, IL **Apr. 4 - 7, 2024**
The Waves of US-China Technology Decoupling and Its Implication on MNCs: Using VAR Time Series Approach

2024 SPSA Annual Conference — New Orleans, LA **Jan. 10 - 14, 2024**
Navigating the Waves of US-China Decoupling: A Comparative Analysis of Trade and Investment in Technology and Non-Technology Sectors

2023 ISDSA Meeting — Shanghai, China **Jul. 4 - 6, 2023**
China's COVID Lockdown Policy and Trade with US: A Deep Learning Time Series Approach

2022 APSA Annual Meeting & Exhibition — Montreal, Quebec, Canada **Sept., 2022**
Framing 2018 US-China Trade War during the Trump and Biden Eras (Accepted)

2022 ISDSA Meeting — Notre Dame, IN, USA. **May 31 - Jun. 1, 2022**
Modeling US-China Trade Relations: A Time Series Machine Learning Approach Using MNC Stock Data

Publications

Hanson-DeFusco, J., Shi, M., Du, Z. et al. 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. *Global Health* 19, 96 (2023). https://doi.org/10.1186/s12992-023-00997-8/

Luhui Yang, Min Shi. 2023. An Analysis of the Motivation of the Abe Cabinet's Policy Evolution and Adjustment towards China. *Journal of China's Neighboring Diplomacy.* Vol.6, No.2, 43-68.

Luhui Yang, Min Shi. 2019. China Policy Adjustment or Changes by the Abe Administrations and Its Impacts. *Peace and Development.* No.3, pp.66-84.

Work Experiences

Reframe Data Services *Data Scientist* **Jan. 2025 – Present**
↪ North Bethesda, MD / Remote

- **Summary:** Full-stack data professional building ML models, analytics dashboards, ETL pipelines, and search-based data modeling solutions to transform raw data into feature-rich outputs.
- Developed and deployed cloud-hosted interactive dashboards, embedding them into front-end applications for real-time data visualization. *Tools: Python, Streamlit, Dash, Heroku, AWS, React, JavaScript*
- Led development of ML/DL models (SVM, CNN, RNN, LSTM) for speaker identification, achieving 80-90% accuracy. *Tools: sklearn, PyTorch, TensorFlow, ML, DL, Transfer Learning*
- Built and optimized real-time and batch data pipelines, integrating OpenSearch and Elasticsearch for scalable data modeling and retrieval. *Tools: OpenSearch, Elasticsearch, AWS Glue, Lambda, Spark, Python, SQL*
- Automated and deployed scalable AI-driven data pipelines for content generation, optimizing efficiency and reliability. *Tools: AWS Glue, Lambda, Streamlit, LLM, Python*
- Worked with engineers and researchers to enhance AI-driven analytics, data infrastructure, and machine learning pipelines.

The Sunwater Institute *Data Engineer* **Jun. 2024 – Oct. 2024**
↪ North Bethesda, MD / Remote

- **Summary:** Designed and optimized scalable data pipelines, big data workflows, leveraging Spark, Hive, AWS for high-performance data processing.
- Built and optimized scalable ETL pipelines for large-scale data ingestion and processing, enabling real-time and batch workflows. *Tools: AWS S3, Glue, Lambda, Spark, Python (Boto3)*
- Developed high-performance data workflows using Apache Spark and Hive on AWS EMR for efficient querying and transformation of big data. *Tools: Spark, Hive, AWS EMR, SQL, Pandas*
- Automated speech-to-text and speaker identification pipelines, improving processing efficiency and achieving 90%+ accuracy. *Tools: PyTorch, CNN, Transfer Learning, AWS Transcribe, Textract*
- Engineered and maintained scalable data infrastructure, integrating CI/CD and monitoring for reliability. *Tools: Airflow, Streamlit, Python, GitHub, AWS EC2, Jenkin*

The University of Texas at Dallas *Data Analyst & Research Assistant* **May 2020 – May 2024**
↪ Richardson, TX / Part-time

- **Summary:** Took responsibility for data manipulation and model building for 10+ global health and policy analytics projects.
- Managed data collection in diverse methods including Qualtrics surveys and web scraping using R and Python.
- Developed 20+ robust statistical models (multi-variable and fixed-effect regression, difference-in-difference, time-series) combined ML models and NLP skills to support correlation and causal inference in research.
- Led a team of five junior assistants, ensuring collaboration and timely project completion and publication.

Lucion Technology Corp., Ltd. *Marketing Data Analyst* **July 2017 – Aug. 2017**
↪ Jinan, CN / Intern

- **Summary:** Served as a Data Analyst Intern responsible for data management, data visualization, and business analysis.

- Improved the efficiency of data extraction by 40% through data optimization in MySQL.
- Employed Microsoft Visio to visualize intricate network structures and aided in product comprehension.
- Produced Business Intelligence (BI) reports, offering insights based on user structures and competitor analysis.

Data Science & ML & Backend Projects

Twitter Clone: High-throughput Social Media Backend ⁻ Python, Django **May 2024 - Present**

- **Summary:** Developing the backend for a social media platform using Django (Python) with HBase, MySQL, and Redis.
- Maximized query efficiency by storing objects with HBase, MySQL & Amazon S3 based on query complexity.
- Addressed N+1 slow query issues by implementing Redis caching and denormalization.
- Integrated Celery and RabbitMQ to establish asynchronous workers with varying priority levels.
- Implemented a push model for distributing news feeds to followers efficiently.
- Optimized memory and resource allocation using recursive small batches of asynchronous tasks.

Strategic Analytics and Planning Report on University-Level Trends **May 2024 - Jun. 2024**

- **Summary:** Conducted a five-year enrollment and graduation trend analysis and provided strategic planning for one Public University.
- Analyzed 17,000+ undergraduate records to identify trends in enrollment and graduation rates by demographics, specifically gender, ethnicity, first generation status, and Pell recipient, supporting strategic decision-making.
- Leveraged multiple data sources to fill gaps, ensuring a thorough report for institutional research and reporting.
- Presented to stakeholders, highlighting key insights to guide strategic planning and continuous improvement.

Cross-Cultural Study of University-Level Survey Data **May 2024 - May 2024**

- **Summary:** Collaborated on a cross-cultural study examining the link between corruption perceptions and academic integrity among university students in Mexico and the U.S., focusing on data collection, statistical modeling, and reporting.
- Collected and cleaned survey data from 535 students across two countries, ensuring accuracy for analysis.
- Developed statistical models to examine the impact of corruption perceptions on academic dishonesty.
- Authored detailed reports, presenting findings and contributing to the study's insights.

US Top 4 Airlines Financial Performance Analytics **Jan. 2024 - May 2024**

- **Summary:** Analyzed 20 years of airline data, identified strategic trends and turning points, and recommended business models.
- Analyzed financial data from a 20-year dataset of over 10,000 rows, covering net income, revenue, and expenses across the US airline industry. This deep dive provided insights into long-term financial trends and shifts.
- Conducted financial performance analytics for the top 4 airlines, identifying key turning points related to major events, alliances, and partnerships over the period.
- Assessed operational trends and competitive positioning of each airline, deriving specific business model recommendations based on a two-decade comparison with competitors.

AI-Powered Payment Service Chatbot for Enhanced Customer Engagement **Aug. 2023 - Dec. 2023**

- **Summary:** Developed and deployed an AI-driven chatbot using Python and MySQL, leveraging advanced NLP techniques to enhance customer engagement for Onyx CenterSource.
- Leveraged NLP and MySQL for analyzing and querying an extensive database containing over 10 million entries.
- Improved response efficiency by 25% and achieved 99% accuracy using the XGBoost model.
- Enhanced user engagement, boosting customer satisfaction and strengthening the company's brand image.

Kaggle Plant Pathology Competition: Leveraging Deep Learning CNNs **Nov. 2023 - Dec. 2023**

- **Summary:** Implemented deep learning models using Python and PyTorch to enhance disease identification accuracy in crops.
- Applied transfer learning on CNNs with 13,042 images across 12 categories, significantly improving accuracy.
- Performed image augmentation techniques (rotation, flipping, zooming, noise injection) to enhance data.
- Fine-tuned ConvNext DL models, achieving 86.8% accuracy and securing a Top 3 ranking in the competition.

Forecasting Stock Prices Through NLP Examination of Newspaper Articles **May 2023 - Dec. 2023**

- **Summary:** Developed automated web scraping, applied NLP techniques to analyze WSJ articles, and improved S&P 500 prediction accuracy.
- Developed automated web scraping for 7,000+ WSJ articles, increasing data acquisition efficiency by 30%.
- Employed various vectorizers for WSJ article analysis, such as Tfidf Vectorizer, n-grams Count Vectorizer, etc.
- Utilized Naïve Bayes and Random Forest models, enhancing S&P 500 prediction accuracy by 12%.

Analysis of the Effect of COVID-19 on US Trade and US Firms **May 2023 - Jul. 2023**

- **Summary:** Built regression and machine learning models for causal analysis, and presented findings at the 2023 Applied Data Science International Conference, earning recognition.
- Synthesized data and created fixed-effect regression models to identify correlations and causal mechanisms.
- Developed and Implemented machine learning and deep learning models to conduct counterfactual analysis.
- Presented research at the 2023 Applied Data Science International Conference to 200 professionals, receiving recognition for clarity and actionable insights.

Analytical Insights and Marketing Strategy Guidance for a Food Company **Feb. 2023 - May 2023**

- **Summary:** Cleaned 1.3 million data records, built interactive Tableau dashboards, and improved forecasting accuracy by 15%.
- Handled data cleaning over 1.3 million raw data records using Python, ensuring data quality and accuracy.
- Developed interactive dashboards in Tableau, enhancing data accessibility and supporting business analytics.
- Employed SAS to construct regression and time series models, leading to a 15% increase in forecasting accuracy.

Optimizing Big Data Risk Analysis for a Company with Hadoop and Tableau **Aug. 2022 - Nov. 2022**

- **Summary:** Engineered data visualization dashboards using Tableau, linked to Hadoop, for business risk analysis.
- Processed and analyzed geospatial data with Hadoop, Hive, Impala and Spark, reducing processing time by 40%.
- Developed Tableau visualizations linked to Hadoop and built interactive dashboards for business matrix analysis.
- Conducted linear regression and multivariate analysis, contributing to predictive accuracy by 15%.

Payroll Management System Database Design via MySQL **Jun. 2022 - Aug. 2022**

- **Summary:** Designed and implemented a payroll management database in MySQL, leading a team of five; developed automated functions, procedures, and triggers, and optimized ETL processes and queries.
- Led a group of five in conducting business requirements analysis and designing a payroll management database with MySQL consisting of 13 tables.
- Created stored functions, procedures, and triggers to calculate employees' payroll per two weeks, fill in new employee information, and send PTO reminders automatically.
- Performed extract-transform-load, data cleaning, and query optimization.

Modeling U.S.-China Trade War's Effect on US Firms using ML and Time Series **Jan. 2022 - May 2022**

- **Summary:** Analyzed the impact of the US-China trade war on MNCs using ML, sentiment analysis, and GARCH time series models.
- A project aimed at exploring how the US-China trade war affects Multinational Corporations (MNCs) through an ML content analysis of policy changes and a time series GARCH modeling approach using stock data.
- Utilized Pandas, NumPy, Matplotlib & Seaborn in data cleaning, visualization, and transformation.
- Leveraged sentiment analysis to explore how the US frame 2018 US-China trade war
- Applied regression analysis in exploring the causal mechanism between trade war and S&P 500 revenues.
- Built machine learning (ML) models in predicting the profound influence of the trade war on US firms.
- Used time-series GRACH models to evaluate MNCs' revenue & volatility quantified via stock data in Stata.
- Presented at 2022 International Society for Data Science and Analytics Conference.

Content Analysis of News Coverage about US-China Trade War **Aug. 2022 - May 2022**

- **Summary:** Analyzed how news organizations framed the 2018 US-China trade war, using machine learning and time-series analysis on over 500 articles to model sentiment trends.
- Led an analysis on how news organizations frame the 2018 US-China trade war during the 2018-2022 period.
- Leveraged machine learning skills such as top modeling and sentiment analysis to explore a collection of over 500 news articles.

- Implemented time-series analysis and chi-squared test in modeling sentiments change tendencies among news coverage.

COVID-19 Worldwide Cases Synchronous Dashboard using Tableau **Dec. 2021 - Jan. 2022**

- **Summary:** Designed and developed an interactive Tableau dashboard to analyze COVID-19 severity worldwide, uncovering key factors influencing the pandemic's impact across countries.
- Designed a synchronous Tableau dashboard with advanced interactive functions to explore the COVID-19 severity.
- Built a Tableau story to dig into the factors affecting the severity of COVID-19 by country and found out the deep connection between multiple aspects of factors with COVID-19 severity.

Selected Course Work

Data Science	**Data Management**	**Data Modeling**
Deep Learning	Big Data	Predictive Analytics for Data Science
Natural Language Processing	Cloud Computing Fundamentals	Modeling for Business Analytics
Causal Analytics and A/B Testing	Database Fundations for BA	Regression and Multivariate Analysis
Programming for Data Science	Information Management	Applied Data Analytics with Python
ML for Socio-Eco and Geo Data	Data Collection	Applied Regression
Content Analysis using ML	Data Visualization	Introduction to Quantitative Methods
OOP in Python	Digital Consulting Project	Social Science Research Methodology
Data Structure & Algorithm	Practical Practicum Project	Prescriptive Analytics

Technical Skills

Programming	Python, R, SQL, Java, HTML, CSS, JavaScript, Stata, SAS
Tools	Amazon Web Services, Streamlit, Dash, OpenSearch, Tableau, Power BI, R Shiny, LaTeX
AI & ML Modeling	PyTorch, Deep Learning, Machine Learning, NLP, Speech-to-Text, Speaker Identification
Database & Big Data	SQL Server, MySQL, PostgreSQL, AWS RDS, Hadoop, Spark, Hive, Impala, Sqoop, Pig
Automation	Alteryx, Appian, Accelq, Uipath
Certificates	Graduate Certificate in Applied Machine Learning at UTD, Google Data Analytics, AWS Certified Cloud Practitioner, Alteryx Designer Core Certificate, Appian Certified Associate Developer, ACCELQ Automation Engineer
Languages	English, Chinese, Japanese