

Jiajun Cheng

Jiajun.cheng@duke.edu | github.com/FraxinusC

EDUCATION

University of Arizona PhD in Management Information Systems	Aug 2026 - Present
Duke University Master of Science in Economics and Computation	Sep 2024 – Jun 2026
Xi'an Jiaotong-Liverpool University Bachelor of Science in Economics and Finance	Sep 2020 – Jun 2024

ACADEMIC PROJECTS

Bluesky Political Polarity & Dynamic Networks Research Assistant for Dr. Timur Kuran, Department of Economics, Duke University	Jul 2025 – Apr 2026
- Cleaned the full Bluesky corpus covering 4M+ users and 235M posts; built a reusable pipeline (dedup, normalization, time alignment, profiling). - Led LLM prompt engineering; drafted guidelines and produced a 200k-sample labeling set (politicalness/leaning) with manual cross-checks; shipped train/val bundles. - Trained classifiers: DeBERTa for politicalness (acc. 97%); RoBERTa-large for leaning (acc. 82%); delivered inference and evaluation scripts. - Next steps: construct and analyze a temporal political-polarity network (community structure, cross-camp edge share, assortativity, polarization index, visualizations); integrate with an <i>expression–reputation–network coupling</i> model; perform parameter calibration and draft controlled experiments; contribute to results interpretation.	
Emphasis–Insight Decomposition for Clinical Notes: MI-Based Mortality Prediction on MIMIC Research Assistant for Dr. Kan Xu, Department of Information Systems, Arizona State University	Jul 2025 – Jan 2026
- Built the MIMIC cohort and preprocessing pipeline aligning clinical notes with structured vitals/labs; standardized timestamps and generated in-hospital mortality labels. - Designed a Transformer-based mutual information framework to extract salient insights from nurses' notes and GPT-generated contexts, effectively modeling clinical documentation patterns. - Developed a text–tabular hybrid prediction model with interpretability outputs, and constructed a hypergraph model to learn clinic notes relation. - Currently finalizing per-visit <i>emphasis/insight</i> attribution, conducting style–semantics ablations and counterfactual note edits, quantifying temporal information gains, and preparing manuscript and code release.	
A View Across the Skyline: Nowcasting Combining Satellite and Energy Indicators Coauthor with Dr. Bo Jiang, Department of Economics, Xi'an Jiaotong-Liverpool University	Jun 2023 – Jun 2025
- Trained on cloud-free median daytime satellite imagery (2017–2021) for 3,113 U.S. counties, processed with GIS and Google Earth Engine, in conjunction with regional energy, and demography data. - Employed a ResNet-50 convolutional neural network based on PyTorch for satellite image processing. - Integrated the extracted image features with energy and population characteristics using an Attention Feature Fusion (AFF) mechanism, optimizing the data flow for subsequent training. - Achieved an impressive R^2 of 0.95 in a 50-epoch training on California's 58 counties, outperforming existing daylight satellite research by 5% and conventional nighttime imagery studies (0.7). - Manuscript under review, preprint available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5101867.	
Attacks and Defenses in Retrieval-Augmented Generation with Large Language Models Group Member, Duke University	Jan 2025 – May 2025
- Conducted a systematic study of GPT-4-style models' vulnerability to misleading retrieved passages in RAG pipelines, framing six bias-driven attack families (Authority, Length, Multilingual, etc.). - Curated a 200-question TriviaQA subset paired with purpose-built "false-context" documents, and introduced the Accuracy-with-False-Context (Accf) metric to quantify robustness.	

- Benchmark tests showed instruction-hijack attacks slashed Accf to 6.5% under naive prompting, while explicit self-guided confidence reasoning restored performance to 59%.
- Developed an evolutionary attack pipeline that uses defense-side explanations to iteratively refine adversarial documents, exposing “explanation–decision” gaps.

Probabilistic Point Process Prediction: Uncertainty-Aware Temporal Modeling

Oct 2024 – Dec 2024

Group Member, Duke University

- Built on Temporal Point Processes (TPPs)—probabilistic models that capture when and what events occur in continuous time—to reproduce the deterministic RMTTP baseline and design two Gaussian-stochastic variants (RMTTP-2, RMTTP-3) plus a GAN-augmented generator-predictor (GAN-RMTTP).
- Refactored each extension into modular PyTorch layers, enabling Monte-Carlo likelihood evaluation and adversarial pre-training within the original EasyTPP pipeline.
- Cut time-prediction error on the NYC Taxi benchmark from RMSE 0.39 → 0.35 and achieved joint-best 0.21 RMSE / 0.52 Accuracy on long-horizon Taobao clickstreams—an $\approx 10\%$ improvement over deterministic baselines.

Architecture Analysis of U.S. Banks and Financial Institutions

Jun 2023 – Oct 2023

Research Assistant for Dr. Bo Jiang, Department of Economics, Xi'an Jiaotong-Liverpool University

- Visualized corporate architecture of banks/institutions based on FFIEC-provided structure tables.
- Investigated correlations between business network topology modifications and financial regulations.
- Used a python-based depth-first search (DFS) algorithm to traverse the nodes and relationships of the study population, calculating the network architecture of 4080 firms using 217,206 corporate shareholding changes from 1970 to 2021.
- Utilized NetworkX and Pyvis in Python for interactive visualization, creating a dynamic network graph illustrating the corporate relationships of the study companies.

TEACHING EXPERIENCE

Teaching Assistant: Programming for Data Analytics (Fuqua); Fraud Analytics (Fuqua); Topics In Computer Science

PROFESSIONAL EXPERIENCE

Deloitte China

Jul 2023 – Sep 2023

Tax and Business Advisory Services Intern

- Streamlined tax return processes by leveraging Python to automate data entry and calculations for 30 complex tax returns, resulting in improved efficiency and accuracy.
- Played a key role in bid preparation for data analysis and visualization, provided crucial data support for the creation of proposals and made use of economic models to forecast the financial viability of projects, contributing to informed decision-making.
- Conducted 8 in-depth studies on tax issues, utilizing statistical and economic models to analyze the impact of various tax policies on the client's economic activities.

Shenwan Hongyuan Securities Co., Ltd

Jan 2022 – Apr 2022

Bond Underwriting Intern

- Engaged in 5 corporate bond and 2 government debt financing instrument issuance projects, with responsibilities including project management assistance, meeting minutes, and data collation.
- Prepared 6 preliminary reports for corporate due diligence, contributing to the successful completion of bond issuances valued at over RMB 1 million.
- Organized project documents for compliance, collated essential corporate information, and assisted in decision-making by estimating bond bid scores and verifying corporate integrity.
- Developed a Python crawler that automated data collection, reducing manual searching and downloading time by 90% and increasing efficiency for the team.

INTERESTS

Interests: Hiking, Mountaineering, Cat Caregiving, Photography