

Liancheng Gong (Krystal)

CONTACT INFORMATION	College Park MD 20742	lianchenggong0224@gmail.com gonglc@umd.edu
EDUCATION	University of Maryland Doctor of Philosophy in Information Studies Advised by Assistant Professor Julia Mendelsohn and Assistant Professor Wei Ai	College Park, MD August 2025 - present
	University of Pennsylvania Master of Science in Engineering in Data Science, GPA: 3.90/4.00 • Math: Big Data Analytics, Statistics for Data Science • Computer Science: Database and Information Systems, Intro to Software Develop • Artificial Intelligence: Principles of Deep Learning, Artificial Intelligence, Ethical Algorithm Design	Philadelphia, PA August 2022 - May 2024
	New York University Bachelor of Science in Data Science, GPA: 3.81/4.00 • Math: Probability and Statistics, Ordinary Diff Equations, Introduction to Math Modeling • Core Data Science: Data Structures, Machine Learning, Information Visualization, NLP, Algorithms • Human-Centered: Decision Making Under Uncertainty, Human-Centered Data Science	New York City, NY Abu Dhabi, UAE Shanghai, China August 2018 - May 2022
HONORS AND AWARDS	Dean's Fellowship, University of Maryland Magna Cum Laude, New York University Dean's List for Academic Year, New York University	FY26 2022 2019, 2020, 2021, 2022
RESEARCH EXPERIENCE	Computer Science department, Drexel University Researcher working with Assistant Professor Harry Zhang • Structured Large Language Models (LLM) outputs into Planning Domain Definition Language (PDDL) domain and problem files to generate verifiable plans • Iteratively refined outputs using environmental feedback to improve generation accuracy • Categorized recurring errors across trials, experimenting with solutions including memory integration and partial file editing to enhance model robustness	Philadelphia, PA September 2024 - July 2025
	Political Science department, University of Pennsylvania Research Coordinator (full-time) working with Professor Daniel Hopkins • Analyzed polls and exit polls for FiveThirtyEight's 2024 presidential election coverage • Investigated Democratic and Republican advertisement campaigns across Meta, Instagram, and Google; examined topic trends in Russian media outlets, such as anti-government narratives • Scraped and processed 2M+ VKontakte posts from 19 media outlets via HPC, created feature vectors using TF-IDF and BERT embeddings, developed a logistic regression model to identify political posts, increasing the sample's political content from 10% to 50% • Annotated Russian blurbs via OpenAI API, including prompt engineering for annotation and translation, achieving 90% alignment between GPT and human annotations on political features • Classified symbolic rhetoric by generating embeddings and fine-tuning on BERTweet to detect topics like "us vs them" and abstract themes	Philadelphia, PA August 2024 - July 2025
	Computational Social Science Lab, University of Pennsylvania Data Scientist working with Yuxuan Zhang and Regina (Jeanne) Ruane	Philadelphia, PA February 2024 - May 2024

- Designed a master list to annotate and analyze bias in media discussions on topics like Biden's age and NATO, delivering findings to the [dashboard](#)
- Automated **extract, transform, and load (ETL)** processes for streaming Nielsen data, improving data consistency and minimizing manual handling
- Managed an **AWS-hosted database**, ensuring data accuracy and schema consistency; standardized network names across large datasets for enhanced data cohesion and analysis

Graduate Research Assistant

February 2023 - December 2023

Atlas working with [Mark Whiting](#)

- Segmented and mined information from PDFs, leveraging **OpenAI's question-answering API** and context window techniques, e.g. Dynamic Memory Networks, Parallel Context Windows
- Pioneered **LLMs** approach by refining prompts and GPT models, achieving superior comprehension and context alignment
- Sourced a robust dataset of 300+ papers and crafted 300+ precise questions, covering all critical analysis units and elevating the accuracy and depth of the literature analysis

Human Mobility Data for Epidemiology working with [Jorge Barreras Cortes](#)

- Developed **Susceptible-Exposed-Infected-Recovered models** on contact networks utilizing **Poisson Distribution** and **Tensor** to predict virus spread, guiding intervention strategies
- Managed **ETL** of 100GB+ databases via **AWS RDS and EMR**; conducted cross-analysis on POI distribution and mobility indicators, ensuring model's applicability and accuracy
- Optimizing the **DBSCAN** clustering algorithm on population data, providing insights of area-specific visit patterns

PROJECTS

Machine Q&A on Chinese Judicial Reading Comprehension [\[Code\]](#), [\[PDF\]](#)

Developed a Span-Extraction Machine Reading Comprehension (MRC) system

Colorize images with Deep Learning [\[Code\]](#), [\[PDF\]](#)

Compared ResNet, GAN, VAE to colorize grayscale images

Customer Lifecycle, NYU X HSBC [\[Code\]](#), [\[PDF\]](#)

Segmented customers, predicted inactivity, and provided targeted recommendations for HSBC

A Model to Predict the Maximum Car Speed in the Rain [\[Code\]](#), [\[PDF\]](#)

Quantified the visibility of the driver and calculate maximum speed under safe speed consideration

TEACHING EXPERIENCE

Fife Academy Coding Club, University of Pennsylvania

Philadelphia, PA

Coding Instructor

October 2022 - May 2023

- Developed and led a modular coding curriculum for 100+ K-12 students, with a 100% positive feedback rate, delivering weekly lectures, hands-on labs, and individual feedback

WORKING EXPERIENCE

Tencent Technology (Shenzhen) Co., Ltd

Shanghai, China

Data Scientist, Platform and Content Group

May 2021 - July 2021

- Enhanced user segmentation and ad CTR by 5% through analysis of 100M+ user data (Python, MySQL); developed ARIMA (0.073 RMSE) for activity forecasting and used NLP on search logs, increasing retention by 4% via A/B testing

Shanghai Urban Real Estate Appraisal Co., Ltd

Shanghai, China

Data Analyst, Data R&D department

December 2020 - April 2021

- Developed a predictive model (Regression, Random Forest, XGBoost) with 80% R² for rental pricing, supported by literature review, survey (300+ responses), and a 2000+ entry database from field visits and web scraping

- COMPUTER SKILLS
- Programming Languages: Python, R, SQL, Java
 - Frameworks and Libraries: NumPy, Pandas, SciPy, Scikit-Learn, PySpark, Seaborn, Matplotlib, PyTorch, TensorFlow
 - Tools and platforms: AWS (RDS/Sagemaker/S3/EMR), MySQL, MongoDB, Tableau, PowerBI, Git, Linux, Stata, Matlab, ETL