

Ann Lian

Washington, DC | (661) 714-6747 | al1698@georgetown.edu | <https://www.linkedin.com/in/annlian>

Data analyst skilled in Python, R, and SQL, with experience in data processing, automation, and public opinion research.

Strong in data cleaning, troubleshooting, and delivering actionable insights.

EDUCATION

M.S. in Data Science for Public Policy, Georgetown University, Washington, DC 05/2025

- Courses: Advanced Statistics, Strategic Policy, Microeconomics, Public Management, Foreign Policy, Massive Data

B.S. in Political Science - Data Analytics, University of California - San Diego, San Diego, CA 06/2023

- Graduated with High Honors
- Courses: Text Analysis, Conflicts in Technology, International Relations, Political Economy, Indo-Pacific Relations
- Honors Thesis: Applied Machine Learning and NLP to analyze electoral disinformation patterns | [Link](#)

PROFESSIONAL EXPERIENCE

Data Analyst - Social Media, Massive Data Institute, Washington, DC 09/2024 - Present

- Enhance misinformation detection and risk assessment workflows by processing and analyzing 100K+ posts from platforms like X, Reddit, and TikTok using **Python, SQL**, and Google Custom Search **API**
- Developed automated **ETL pipelines** using **Python** to streamline data collection and cleaning, increasing extraction efficiency by 30% and supporting real-time monitoring dashboards
- Improved model precision by 15% through application of **supervised learning (Logistic Regression, Random Forest)** and behavioral modeling techniques to identify disinformation patterns and user engagement strategies
- Designed interactive dashboards using **Plotly Dash** and **Tableau**; drafted client-facing policy briefs summarizing data trends, key risks, and actionable recommendations for government and civil society partners

Researcher, University of California - San Diego, San Diego, CA 01/2022 - 03/2023

- Led cross-disciplinary research integrating social media analytics and geopolitical analysis, applying **unsupervised learning, sentiment analysis**, and **keyword co-occurrence** methods to evaluate user behavior and narrative targeting
- Analyzed 160K+ multilingual posts (English, Mandarin, Russian) using **Python (scikit-learn)** and **R**, uncovering a 13% surge in engagement with politically charged content during key international events
- Boosted model performance by 18% through extensive data preprocessing (**language normalization, lemmatization**) and regression **optimization** using **cross-validation** and **hyperparameter tuning**

Deloitte Leadership Fellow, McCourt Public Policy School, Washington, DC 05/2024 – 05/2025

- Spearheaded leadership development and onboarding initiatives for incoming graduate students, strengthening team cohesion through peer coaching, conflict resolution training, and inclusive leadership workshops
- Improved onboarding by creating a structured orientation tailored to stakeholder needs and operational timelines

Senior Legal Intern, California State Senator's Office, San Diego, CA 02/2022 - 01/2023

- Conducted legislative research and quantitative policy analysis on economic and political issues
- Managed sensitive policy and fiscal data in full compliance with confidentiality protocols

PROJECTS

Geospatial Analysis of Natural Resources and Conflicts | R, GIS | [Link](#) 01/2025

- Integrated 10,000+ geospatial data to analyze resource distribution and regional conflict risks, supporting risk assessments
- Applied spatial econometric models to identify statistically significant conflict clusters and regional autocorrelation
- Developed interactive dashboards and map visualizations using **R** and **GIS** tools to present findings

Mapping Media Narratives | Python, APIs 12/2024

- Developed a hybrid framework combining network and text analysis to evaluate media narratives and detect bias
- Automated large-scale web scraping and implemented validation processes, improving accuracy and efficiency

Impact of Cyber Attacks on International Alliances | Python, TensorFlow, Scikit-learn, Matplotlib, Seaborn 05/2024

- Merged 1,000+ cyber incidents with military expenditure data to evaluate global cybersecurity risks and defense responses
- Built and validated predictive models to forecast military spending trends, achieving 98.45% accuracy
- Conducted correlation and time-series analyses to assess alliance stability and strategic defense decisions

RELEVANT SKILLS

Technical: Python, SQL, R, Stata, Tableau, Excel (VBA), GitHub, APIs, Shiny, Power BI, GIS, AWS

Analysis: Regression, Classification, Econometrics, Risk Modeling, Customer Segmentation, Litigation Research

Languages: English (fluent), Mandarin Chinese (native), Taiwanese (native), Japanese (intermediate)