

# Tong Zhu

919-694-7177 | toz015@ucla.edu | www.linkedin.com/in/tongzhu0223

## Summary

With 2-years working experience in data analysis, **Machine Learning** and **A/B testing**. Solid programming skills in **Python, JAVA, SQL, R**

## Education

### University of California, Los Angeles

Doctor of Philosophy (PhD) in Biostatistics

- **Research Area:** LLM Agents Mechanism

Los Angeles, CA

Sep 2024 - Current

### Northeastern University

Master of Science in Computer Science

- **Courses:** Algorithm, Distributed Database

San Jose, CA

Feb 2023 - Sep 2024

### University of North Carolina at Chapel Hill

Master of Science in Statistics and Operations Research

- **Courses:** Applied Statistics, Machine Learning, Time Series Forecasting

Chapel Hill, NC

Aug 2018 - July 2020

## Work Experience

### Amazon

Software Engineer Intern

- Developed features for Amazon B2B on **AWS** to automate event-driven applications using **EventBridge**.
- Enhanced the *Visibility Service* using **Java**, **JavaScript**, and **TypeScript**, improving real-time monitoring of critical business transactions.
- Wrote and executed integration tests for outbound services and events, enabling smooth, error-free deployments.
- Designed and validated **CloudWatch** dashboards to monitor key service performance metrics.

Los Angeles, CA

Jun 2024 - Aug 2024

### ByteDance Ltd.

Data Scientist

- Collaborated with cross function team to deploy dynamic subscription tool to improve customer experience in platform
- Verified product feasibility and deployed **Xgboost** model to select the important indicators for customers to help design product features
- Applied **interrupted time series** to estimate the potential revenue impact, and evaluate the risk of restricting creator's quote strategy in advance
- Conducted **attribution analysis** to evaluate marketing campaign performance which provided powerful evidence to spur on product marketing

Shanghai, China

Feb 2021 - Aug 2022

### TouchSuite

Machine Learning Intern

- Queried and cleaned terabyte-sized order data from **Azure SQL** using pyodbc
- Conducted **online analytical processing (OLAP)** to display critical sales performance from different dimensions
- Developed item-based approaches to handle cold-start problems and tuned the model hyper-parameters through **Spark ML cross-evaluation toolbox** which reduced root mean square errors by 10%

Boca Raton, FL

Jun 2020 - Aug 2020

## Publications

### NeurIPS 2025

Incentivizing Truthful Language Models via Peer Elicitation Games

- Baiting Chen<sup>\*</sup>, **Tong Zhu<sup>\*</sup>**, Jiale Han, Lexin Li, Gang Li, Xiaowu Dai. *Incentivizing Truthful Language Models via Peer Elicitation Games.* .

Sep 2025

## Projects

### Natural Language Processing and Topic Modeling on Customer feedback

University of North Carolina at Chapel Hill

- Clustered customer feedback into different types and discovered the latent semantic structures with **Python**.
- Preprocessed feedback text by **tokenization**, **steming**, stopwords removing and extracted features by **Term Frequency-Inverse Document Frequency (TF-IDF)** approach
- Trained unsupervised learning models of **K-means clustering** and **Latent Dirichlet analysis** and Identified latent topics and keywords of each feedback text for clustering by **Python Sklearn**

Chapel Hill, NC

Jan 2019 - March 2019

## Skills

**Programming Language** Python (Pandas, Seaborn, NumPy, Scikit-learn), R, Java, and SQL

**Tools and Platforms** MySQL, Tableau, AWS, and Spark