Benny Xinhao Jiang

bennyjxh@gmail.com | 510-207-4905 | linkedin.com/in/benny-jiang | bennyoooo.github.io

EDUCATION

Carnegie Mellon University — Pittsburgh, PA
M.S. in Computational Data Science, Analytics Concentration — May 2021 - May 2023
- **QPA: 4.01/4.00**
- Relevant Coursework: Cloud Computing, Search Engine, Deep Learning, Advanced Natural Language Processing, Machine Learning, Large-Scale Multimedia Analysis.

University of California, Berkeley — Berkeley, CA
B.A. in Computer Science & Data Science, emphasis in Applied Mathematics & Modeling — Aug 2017 - May 2021
- **GPA: 3.93/4.00** (High Distinction in General Scholarship)
- Relevant Coursework: Database Systems, Operating Systems, Deep Reinforcement Learning, Parallel Programming.

SKILLS

Tools & Technologies: Spark(ETL, SparkML), Cloud Computing(AWS, Azure, GCP, Kubernetes), Docker, MySQL.

Programming Languages: Python(PyTorch, Scikit-learn, NumPy, SciPy, Pandas), Java, C++, Javascript.

PROFESSIONAL EXPERIENCES

Meta — New York, NY
Software Engineer Intern — May 2021 - Aug 2021
- Worked on machine learning modeling and signal serving on **the Instagram Reels recommender system** team.
- **Data Engineering:** assembled data pipelines that compute engagement time for ~20 million posts every day.
- **Modeling:** Built and trained multitask multi-label deep learning models that predict the time-sensitivity of media.
- **Signal Serving:** Deployed trained models as public features and onboarded reliable real-time signals.
- **Value Model Optimization:** Designed and implemented value model logic that promotes fresh and timely media in the ranking system, increasing viewer watch time by over 2%.
- Launched language mismatch filter in indexing system, which improves viewer engagement metrics by over 3%.

Tencent — Shenzhen, China
Software Engineer Intern, Financial Technologies — May 2021 - Aug 2021
- Collaborated with a team of 15 to improve the transaction and user management of digital currency platforms.
- Built an RPC server using C++ that enabled customers to forge secure payment channels with merchants.
- Upgraded an RPC framework that supported multi-process reverse proxy, and load balancing of servers.

Hulu — Santa Monica, CA
Software Developer Intern — Jun 2020 - Aug 2020
- Created a Flask server reducing complicated procedures of cross-platform troubleshooting into one searching action.
- Constructed an integration platform that monitored the status of seven micro-services, saving time of engineers across five teams for inspecting and debugging the servers.
- Led a project team of 3 to create a Chrome extension with JavaScript that let users collect and share video clips in TV shows and won a special award in the internal corporate hackathon.

BizSeer — Beijing, China
Research Data Scientist Intern — Jun 2019 - Aug 2019
- Co-authored and published a paper: "Diagnosing root causes of intermittent slow queries in cloud databases" in Proceedings of the VLDB Endowment 13, no. 8, April 2020.
- Proposed a clustering algorithm combined with case-based reasoning models to create an active learning pipeline that queried the labels of representative data points from database experts, efficiently easing labeling overhead.
- Developed a real-time anomaly detector that extracted anomalies on time series, with over 97 F1 score.

PROJECTS

Cloud Computing — Feb 2022 - May 2022
- Developed a high-throughput fault-tolerant web server with a web tier and a storage tier micro-services, containerized with Docker, deployed on AWS clusters with Elastic Load Balancer, and orchestrated with Kubernetes.
- **Web Tier:** Constructed a QR code processor, a blockchain validator, and a Twitter recommender system micro-services with stable throughputs of over 185k ,100k, and 30k QPS, respectively within an AWS budget of $0.7/s.
- **Data Processing:** Implemented ETL on 1TB of raw Twitter data with Spark on Azure HDInsight clusters.
- **Storage Tier:** Loaded the data into MySQL databases on AWS RDS; optimized the database schema and applied the MyISAM storage engine and read-intensive loads.
- **Automated** the deployment of containerized micro-services to Kubernetes using Helm and Terraform.

QryEval Search Engine — Jan 2022 - April 2022
- Built a search engine evaluating queries with ranked/unranked boolean, BM25, or Indri retrieval models.
- Developed Pseudo Relevance Feedback for query expansion and explicit diversification methods xQuAD and PM2.
- Set up learning-to-rank feature collection and training funnels and trained SVM LeToR models.