

# Dongxiao Han

seeking 2022 Summer SWE Backend intern position

Mobile: (412) 214-2378

Email: [dongxiah@andrew.cmu.edu](mailto:dongxiah@andrew.cmu.edu)

LinkedIn: <https://www.linkedin.com/in/dongxiao-h-585414b1/>

Location: Pittsburgh, PA

## EDUCATION

### Carnegie Mellon University

Pittsburgh, PA

Master of Information Systems Management, GPA: 3.7/4.0

Feb 2021-Dec 2022 (Expected)

Selected Coursework: Parallel Computer Architecture and Programming (15-618, Spring 2022), Machine Learning in Production (17-645, Spring 2022), Computer Systems (15-513, Fall 2021), Search Engines (11-642, Spring 2021)

### Shanghai University

Shanghai, China

Bachelor of Science in Information Systems

Sep 2021-Jun 2018

## SKILLS

**Language** (Proficient) Java, Python; (Familiar) C/C++, Scala, JavaScript, SQL, HTML, CSS, R

**Framework / Tool** Spring Boot, Django, Flask, React, Spark, PyTorch, Docker, Git, Nginx

**Database** MySQL, PostgreSQL, Redis, MongoDB

**Cloud / OS** AWS (EC2), AliCloud (ECS, OSS), Linux

## WORK EXPERIENCE

### Software Engineer Intern

Shanghai, China

SAS Institute

Apr 2021-Jul 2021

- Built a data integration web platform from scratch where clients can connect with their databases, clean data, and make model prediction experiments online with friendly UI without coding.
  - **Increased monthly revenue contribution by 8% for 12 clients** as a value-added service bundled with SAS JMP.
  - Optimized performance of distributed matrix algorithm based on MapReduce, **decreasing running time by 80%**.
  - Implemented spreadsheet editor similar to Excel, supporting caching and **decreasing waiting time by 23%**.
  - Used React and JavaScript for frontend, Flask for backend, Redis and MySQL for storage, Docker for deployment.
- Launched a crawling system to collect online data for industry insight data product.
  - Optimized crawling techniques and **increased crawl rate by 16%, achieving 98% data integrity**.
  - Collected over 1 million daily data in Python and store data in PostgreSQL.

### Data Scientist (Software Engineering Track)

Shanghai, China

Alibaba Group (China Business Network)

Jul 2020-Mar 2021

- Built a deep learning image retrieval system to retrieve similar clothing based on uploaded images.
  - **Led a 3-member team** (frontend engineer\*1, UI designer\*1), responsible for model training and backend development.
  - Embedded the system to a data product and **increased user engagement by 23%**.
  - Customized data structure and caching techniques, which **achieved 6.7X speedup** for image retrieval from 190K images.
  - Used React for frontend, Django for backend, Redis for cache, PyTorch for model training.
- Built an SMS promotion platform where can send advertising messages and make AB Testing experiments.
  - Optimized sending speed with multithread program, **achieving average 3700 QPS on a single server**.
  - Implemented interactive AB testing monitor dashboard, which can assign experiment groups and make statistical analysis.

## PROJECTS

### Dynamic Memory Allocator

Carnegie Mellon University, Fall 2021

- Utilized C language to implement methods of block assignment on the heap similar to malloc.
- Implement segregated free list for block coalescing and splitting, achieving 61% utilization rate, 15 million OPS, ranking 11/363.

### Recommendation System with Online and Offline Storage

Carnegie Mellon University, Spring 2021

- Implemented feature engineering, embeddings, and collaborative filtering algorithm on 1M users with Scala Spark.
- Utilized JavaScript for frontend, Spring Boot for backend, Redis for online storage, and Hadoop for offline storage.

### Lucene-based Search Engine System

Carnegie Mellon University, Spring 2021

- Used Java to build search algorithms, including exact match model (Ranked Boolean) and best match model (BM25, Indri).
- Reproduced Deep Relevance Matching Model and trained neural network using pairwise method, achieving 0.82 NDCG@10.