

# TENGDA WANG

tengdaw@cs.cmu.edu | +1 (412) 214-2825 | [Linkedin](#)

## Education

### Carnegie Mellon University

Pittsburgh, PA

Master of Computational Data Science (System Track), **4.0/4.0** GPA

Aug 2023 - Nov 2024 (Expected)

- **Key Courses:** Intro to Computer Systems, Database Systems, Distributed Systems, Advanced Cloud Computing

### National University of Singapore

Singapore

B.Sc. (Honors) in Business Analytics, **4.87/5.0** GPA

Aug 2017 - May 2021

- **Awards:** Winner of [IMDA excellence Prize](#) (most outstanding graduating student); Valedictorian for Graduation Ceremony; Dean's List for 3 semesters (top 5% of the cohort); Singapore Ministry of Education Tuition Grant
- **Activities:** Teaching Assistant for 2 Data Structures & Algorithms modules and 1 Machine Learning module
- **Key Courses:** Natural Language Processing, Web Development, Software Engineering, Computer Organization, Network

## Experience

### Shopee

Singapore

Software Engineer - Machine Learning, Search Team

July 2021 - May 2023

- **Query-Category Relevance:** Improved search quality for million users by boosting items that are relevant against the search query. Achieved 92.4% offline training AUC, increased CTR by 4.21% and decreased bad case rate by 20.21% online.
- Architected, implemented, and maintained an end-to-end distributed pipeline consisting of a large data warehouse, an ETL feature-extraction pipeline, a model training and inference module, and a deployment layer utilizing Redis Cache, that handles up to 100 TB data size and thousands QPS.
- Pre-trained monolingual BERT models using a masked language task on item descriptions in 8 different languages, which improved performance of downstream tasks (NER, query rewrite etc) in both feature-based and fine-tuning fashion.
- Worked collaboratively with 5+ Product Managers and Product Ops across 8 countries to generate 10 million rows of human-labeled data for model training, which significantly increased model offline metrics by 44.2%.
- Productionalized a modified K-Fold cross-validation algorithm to detect labeling errors in human-annotated data and enabled weighted training based on confidence measures, which improved offline AUC by 1.2% and saved at least 10 hours of labor work per week.

### Bank of America Merrill Lynch

Singapore

Software Engineer Intern, Global Markets Tech Team

Jun 2020 - Aug 2020

- Designed workflows and created multiple full-stack web applications including frontend, backend, and unit-testing to help clients manage portfolios. Worked closely with the product side to ensure a well-designed UI/UX.

## Projects

- [Distributed Proxy](#) (2024): Designed and coded a distributed proxy server that supports whole-file caching and LRU eviction. The proxy uses Java RMI as the underlying RPC protocol, and leverages check-on-use techniques to ensure cache consistency in open-close session semantics similar to the Andrew File System (AFS). (**Java**)
- [Bustub](#) (2023): Extended the functionality of a relational database management system by implementing a buffer pool manager with LRU-K eviction policy, an extendible hash index and a B+ tree index supporting concurrent operations, a query execution engine, and a multi-version concurrency control protocol for transactions. (**C++**)
- [15513 Labs](#) (2023), Malloc Lab: wrote a segregated-list version of dynamic memory allocator that supports malloc, free and realloc to maximize utilization and throughput; Proxy Lab: implemented a multi-threaded proxy server utilizing caching; Shell Lab: developed a process-based unix shell that supports job control and I/O redirection. (**C**)
- [Bachelor Thesis](#) (2021), Studied the use of neural abstractive summarization techniques to automatically generate hospital discharge summaries conditioned on nursing notes in electronic health records (**Pytorch, Tensorflow**)
- [NUS Internship Search](#) (2019), Developed and deployed a visualization web application to provide graphical and analytical insights for students to find internships (**Vue.js, HTML/CSS, Node.JS, Firebase, Highcharts, Travis CI, Heroku**)

## Skills

- **Programming Languages:** C, C++, Scala, Java, Python, Go, R, Javascript (Angular/Vue/React), HTML/CSS, Bash
- **Data:** PostgreSQL, Redis, Hive, Spark, MapReduce, HDFS, Airflow
- **Data Science:** Scikit-Learn, Pytorch, Tensorflow, Tableau
- **Other:** Docker, Kubernetes, AWS, Terraform, REST, Protobuf, Linux, Git