

# TENGDA WANG

tengdaw@cs.cmu.edu | (412) 214-2825 | <https://cs.cmu.edu/~tengdaw> | [Linkedin](#)

## Education

### Carnegie Mellon University

Pittsburgh, PA

Master of Computational Data Science (System Concentration), **3.89/4.0** GPA

Dec 2024

- **Key Courses:** Database Systems, Distributed Systems, Advanced Cloud Computing, Deep Learning Systems, Parallel Architecture And Programming, Intro Computer Systems, Search Engines, Machine Learning, Natural Language Processing

### National University of Singapore

Singapore

Bachelor of Science (Honors) in Business Analytics, **4.87/5.0** GPA

May 2021

## Skills

**Programming Languages:** Python, Java, Scala, C, C++, Scala, Go, Javascript (Angular/Vue), Bash, SQL

**Data:** Apache Spark, Apache Kafka, Hadoop MapReduce, Apache Airflow, Apache Hive, HDFS, Redis

**Machine Learning/Data Science:** PyTorch, TensorFlow, Scikit-Learn, Keras, Tableau

**DevOps/Cloud:** Docker, Kubernetes, Terraform, AWS (EC2, S3, Lambda, CWL, SageMaker), GCP, Protobuf, Linux, Git

## Professional Experience

### TikTok

Bellevue, WA

Machine Learning Engineer Intern, E-commerce Recommendation Team

Jun 2024 - Aug 2024

- Developed and iterated on multiple **collaborative-filtering** based algorithms in **HiveSQL** and **C++** for efficient product **recall** in a large-scale recommendation system (TikTok Shop).
- Implemented and trained two-tower **embedding recall models** optimizing click-through and conversion, with a focus on feature engineering, model structure, and data pipeline. Models were deployed online and brought positive GMV impact.

### Shopee

Singapore

Software Engineer in Machine Learning, Search Team

Sept 2021 - Apr 2023

- **Query-Category Relevance:** Improved **search quality** for million users by boosting relevant items in the search result. Achieved 92.4% offline training AUC, increased CTR by 3.71% (A/B test) and decreased bad case rate by 20.21% online.
- Architected, implemented, and maintained an **end-to-end distributed pipeline** made up 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 with 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.
- Collaborated with 5+ Product Managers and Product Ops across 8 countries to generate 10 million rows of human-labeled data for model training, and increased model offline metrics by 44.2%.

### Bank of America Merrill Lynch

Singapore

Software Engineer Intern, Global Markets Tech Team

Jun 2020 - Aug 2020

- Formulated workflows and created multiple **full-stack web applications** including frontend (**AngularJS**), backend (**Scala**), and unit-testing to help clients manage portfolios. Worked closely with product teams to ensure a user-friendly **UI/UX**.

## Selected Projects

- [BusTub](#) (2024): Extended the functionality of a RDBMS 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++**)
- [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**)
- [System Labs](#) (2023), Malloc Lab: wrote a segregated-list version of dynamic memory allocator supporting malloc and free to maximize memory utilization and throughput; Shell Lab: created a process-based unix shell with job control , I/O redirection, and signal handling. (**C**)
- [Bachelor Thesis](#) (2021), Studied **neural abstractive summarization** techniques (e.g. **transformers**, **seq2seq models**) to automatically generate hospital discharge summaries 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**)