Tianyu Zhang
tianyuz.com
Email : tianyuz2@andrew.cmu.edu / Mobile : (650) 505-8722
LinkedIn : johnzhangty

Education

- **Carnegie Mellon University School of Computer Science** Pittsburgh, PA
  B.S. in Artificial Intelligence; 5th-Yr M.S. in Machine Learning. GPA: 3.92/4.00, Dean's List. Expected May 2023
  - **CS**: Database Systems, Distributed Systems, Search Engines, Parallel DS & Algorithms, Software Design
  - **ML**: Deep Learning Systems (PhD), Advanced Deep Learning (PhD), ML with Large Datasets (MS), NLP
  - **Math**: Modern Regression, Intro to Math Finance, Probability & Stats, Multivariate Calculus, Linear Algebra

Work Experience

- **Uber** San Francisco, CA
  Software Engineer Intern *Jun 2021 - Aug 2021*
  - **Dish Recommendation on Uber Eats Home Feed**:
    - Designed and developed a multi-channel framework for dish candidates retrieval in feed service in Go.
    - Implemented a novel random-projection-based embedding retrieval in Java to recall candidates 4x more efficiently than using CVR. Set up training and ingestion workflows to index dishes weekly in the search system.
    - Set up dish-related ETL and dispersal pipelines with Spark and Hive based on order and click data. Implemented other retrieval channels to enrich the candidates set.
    - Prepared feature store pipelines. Trained, tuned, and served models for candidates ranking.
    - Launched the recommended dishes carousel on Uber Eats home feed to 90M global users.
- **ByteDance (TikTok)** Beijing
  Software Engineer Intern *Jun 2020 - Aug 2020*
  - **Live Stream Recommendation with Graph Embedding**:
    - Implemented pipelines to build user-author graphs with billion edges using MapReduce. Served graphs distributedly with millisecond latency using Euler.
    - Implemented neighbor pre-fetching in C++ and reduced internal ML trainer latency by 40% on graphs.
    - Devised graph encoders and end-to-end network architecture with Tensorflow to predict click-through rate.
    - Boosted online user staytime +3.5%, etc. in AB tests and rolled out to 600M TikTok users.
  - **MLOps Systems**:
    - Developed a model health monitor and alert system from scratch in Django with RESTful APIs. Onboarded 100+ online models across 5 products with 50+ internal users. Reduced response time to <1hr.
    - Constructed an analysis pipeline on 300+ features that modifies terabyte model checkpoints distributedly based on analysis result. Saved 35k+ core-hour computing resources than hand-tuning.

Academic Experience

- **TheSys Group, CMU Parallel Data Lab** Pittsburgh, PA
  Research Assistant *Nov 2020 - Present*
  - Researched embedding table fault tolerance in distributed training with Prof. Rashmi K. Vinayak.
  - Experimented with various fault tolerance strategies (replication, checkpointing, and erasure coding) in the open-source training system XDL to understand efficiency tradeoffs.
  - Proposed a novel multi-level approach that utilizes a hybrid fault tolerance strategy to minimize time and memory overhead. Worked on its C++ implementation, benchmarking, and paper drafting.
- **CMU Machine Learning Department** Pittsburgh, PA
  Teaching Assistant for 10-605 Machine Learning with Large Datasets *Feb 2021 - Jun 2021*
  - Designed a major assignment from scratch, with write-ups, tutorial videos, and starter codes. Onboarded 140+ students to ML at scale with Spark and AWS.
  - Wrote exams; led weekly recitations and office hours for 20+ undergraduate and graduate students.

Projects

- **Needle**: (WIP) A PyTorch-like deep learning library with autodiff and GPU acceleration. (C++ , Python)
- **AlpacaHub**: (WIP) An env, data, and model versioning framework for data science workflows. (JS, Python)
- **QASys**: A question generation and answering system on text with rule-based and neural backend. (NLTK, PyTorch)
- **BitcoinMiner**: A failure-recoverable distributed Bitcoin miner with the homegrown Live Sequence Protocol. (Go)
- **Finger**: A tiny-screen-optimized input keyboard with trie and ngram empowered autocompletion. (Java)
- **Pop!**: A crowd-sourcing notification app that allows users to send signals to groups in real-time. (React, Django)

Skills

- **Languages**: Java, Go, C/C++ , Python, SQL, JavaScript, SML/OCaml
- **DevOps**: Spark, Tensorflow/PyTorch, Hive/Presto, gRPC, Docker, HDFS, Kafka, Protobuf, ELK