Yangcheng Gu

Carnegie Mellon University | yangcheg@andrew.cmu.edu | (412) 954-8944

https://www.linkedin.com/in/yangcheng-gu/ | https://ferv3455.github.io/

Education

Carnegie Mellon University, Master of Science in Information Networking GPA 4.00/4.00 | August 2024 – December 2025
- *Courses:* Cloud Computing, Distributed Systems, Database Systems, Algorithms Design and Analysis

Tsinghua University, Bachelor's in Software Engineering GPA: 3.90/4.00 | August 2019 – July 2024
- *Courses:* Computer Architecture, Software Engineering, Data Structures and Algorithms, Artificial Intelligence

Work Experience

Software Engineer (Intern) May 2025 – August 2025

Privacy and Security - Data Lifecycle Management - AI Governance, TikTok Inc.
- Designed and built a full-stack rule management system from scratch, streamlining management of 40+ evolving rules
- Introduced version control and dry-run evaluation using 700+ models, enhancing update auditability and reliability
- Backfilled error data and launched the system to production, enabling real-time downstream consumption of updated rules
- Authored comprehensive design documentation, developer guides and user manuals to ensure long-term maintainability
- Implemented multiple compliance-related tickets to collect AI model information required under current regulations

Algorithm Application Engineer (Intern) June 2023 – August 2023

Risk Control R&D Department, Beijing Jingdong Century Trading Co., Ltd.
- Operated on Apache Hadoop and Spark to compile a dataset and fitted a pre-trained click-farming detection model to it
- Devised and implemented a specialized image similarity algorithm based on statistic methods and feature matching, yielding an average accuracy of over 95% on synthetic datasets and almost 100% on real ones
- Developed a click farming detection pipeline with the proposed algorithm, processing up to 300,000 images per day

Research Projects

Bi-level Optimization for Inductive Transfer Learning July 2023 – July 2024

Computational Biology Department, Carnegie Mellon University
- Proposed and implemented a pipeline for model pre-training and fine-tuning, which assigns sample weights via a neural network optimized with DARTS, boosting its performance in transferring to differently distributed datasets
- Built and executed multi-domain evaluation workflows, which showed a 3% increase in accuracy with a limited dataset
- Integrated the framework with SimCLR self-supervised learning architecture and investigated sample re-weighting using self-supervised deep learning on both natural and medical datasets

Log Data Encoding for Efficient Storage in Apache IoTDB October 2022 – May 2023

School of Software, Tsinghua University
- Introduced and developed advanced character operations and generalized string edit distance as a string encoding scheme
- Implemented character re-weighting based on frequency analysis, slightly improving space efficiency for stored data
- Engineered a log data classification pipeline using clustering to group data by format, reducing required data storage
- Optimized the algorithm temporally with a linear-time cosine distance algorithm for strings based on q-grams

Course Projects

BusTub - Relational Database Management System, Database Systems January 2025 – April 2025
- Developed core components of a relational database management system, including buffer pool manager, efficient database indexes (B+ Tree), query execution/optimization and concurrency control mechanisms
- Skills and tools: C++17, CMake, database system concepts, concurrent programming

Embedded Operating System Development, Embedded Systems September 2024 – November 2024
- Designed and implemented core low-level components of an ARM-based embedded operating system from bare metal, including memory management, device drivers, interrupts, system calls, user-space thread scheduling, synchronization, etc.
- Skills and tools: ARM Assembly Language, C, GDB, system programming, operating systems, serial protocols

MarsOJ Online Coding Competition Platform, Software Engineering September 2022 – January 2023
- Co-led development of the end-to-end web application, driving the design and creation of all user-facing webpages
- Engineered real-time communication mechanism using Socket.IO, enabling low-latency interactions among users
- Skills and tools: JavaScript (Vue.js), Python (Flask), HTML/CSS, WebSocket (Socket.IO), design patterns, software testing

Skills

- **Proficient in:** C/C++, GoLang, Python, JavaScript, Java, HTML/CSS, SQL (MySQL)
- **Familiar with:** Hadoop, Spark, Docker, Kubernetes, AWS, Linux, PyTorch, OpenCV, Qt, Vue.js, LaTeX, Markdown