Pengyu Wang

Website: pengyuwa.ng
Github: github.com/cswpy

Email: pengyu.wang@yale.edu
LinkedIn: pengyu-wang
646-350-9064

EDUCATION

Yale University — New Haven, USA
Master of Science in Computer Science (thesis track) — 09/2023 – 05/2025
Coursework: Principles of Computer System Design, Big Data Systems

New York University — Abu Dhabi, UAE
Bachelor of Science – Computer Science; Cumulative GPA: 3.89, Major GPA: 3.95 — 09/2019 – 05/2023
Coursework: Deep Learning, Database Design, Computer Security, Mining of Massive Dataset, OS

EXPERIENCE

TikTok — San Jose, USA
Software Engineering Intern — 06/2022 – 08/2022
- **Realtime Data Dashboard**: Implemented a streaming data pipeline with **Flink** that monitors real-time ads auditing queues, supporting 450k daily throughput at peak hours for global markets
- Devised a **Lambda architecture** that integrates offline batch sink with online streaming events to serve real-time queries
- Built an extensible pipeline with MQs like **Kafka** and **RocketMQ** and OLAP databases like **ClickHouse** and **Hive**

Nokia — Hangzhou, China
Software Engineering Intern — 01/2021 – 03/2021
- **Automatic Fault-Finding System for Documents**: Developed a website for automatically locating errors in technical documents and displaying them on websites, serving **1000+** engineers worldwide with **18k+** corrections
- Implemented a full CI/CD pipeline and release scheme on development and production servers with **Jenkins** and **GitLab**
- Designed a systematic testing framework using **PyTest** and workflows for unit testing and static checking

combyne — Munich, Germany
Junior Data Scientist — 06/2020 – 12/2020
- Increased first-month retention rates by 2% through running experiments and simplifying the onboarding experience
- Implemented data pipelines using **Jupyter** and **Google BigQuery** to facilitate analytic workflow for other teams

PROJECTS

Linux kernel – out-of-tree Linux kernel module with Rust support: implemented kernel modules that read images from Linux V4L2, transmit to a server for inference, and display results in userspace. Tech: QEMU, Linux, Rust, C, TensorFlow

Grocery on Rails – a cross-platform online grocery: Led the team to implement a full-fledged grocery application with a client mobile app, admin website, and a universal backend. Tech: Python, Flask, Flutter, MongoDB

SimpleDB – implementation of a DBMS: Developed core database components like Tuple, BufferPool, HeapFile, and Transactions management. Tech: Java, SQL, PostgreSQL

RESEARCH

Improving classification performance on Whole-Slide Images — Abu Dhabi, UAE
Course Project — 02/2023 – 05/2023
- Benchmarking SOTA **Multiple-Instance Learning** (MIL) models on Whole-Slide Image classification task
- Implementing a novel deep model with **contrastive self-supervised learning**, knowledge distillation network, and a locality-aware bag classifier

AI in Reducing the Use of Antibiotics in Urinary Tract Infections — Abu Dhabi, UAE
Cranmore Fellowship recipient — 12/2020 – 03/2022
- Leveraged machine learning to predict the presence of UTI with 0.81 AuROC metric
- Trained a **Variational Graph Neural Network** with embedding and GNN classification components

Answering High-Order Optimization Package Queries — Abu Dhabi, UAE
Capstone project — 02/2022 – 05/2023
- Extending relational databases to evaluate Package Queries through SKETCHREFINE heuristics
- Utilizing **K-Dimensional Tree** to partition **10–100** million tuples into groups within **seconds**
- Converting Package Queries into **Integer Linear Programming** problems solved by off-the-shelf optimizers

PUBLICATIONS

Mai, A. L., **Wang, P.**, Abouzied, A., Brucato, M., Haas, P. J., and Meliou, A., "Scaling Package Queries to a Billion Tuples via Hierarchical Partitioning and Customized Optimization", preprint, 2023. arXiv.2307.02860.

Ghosheh, G. O., St John, T. L., **Wang, P.**, Ling, V. N., Orquiola, L. R., Hayat, N., Shamout, F. E., and Almallah, Y. Z. "Development and validation of a parsimonious prediction model for positive urine cultures in outpatient visits", PLOS Digital Health, 2(11) e0000306, 2023. https://doi.org/10.1371/journal.pdig.0000306

COMMUNITY SERVICE

Regional Manager & Head of IT, Education Without Barriers — China
Led 18+ volunteers to teach weekly classes for underprivileged kids in Dandong — 11/2017 – 06/2019