

# 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); GPA: 4.0/4.0* 09/2023 – 05/2025

**New York University** Abu Dhabi, UAE  
*Bachelor of Science – Computer Science; Cumulative GPA: 3.89, Major GPA: 3.95* 09/2019 – 05/2023

## EXPERIENCE

---

**Duolingo** Pittsburgh, USA  
*Operations Engineer Intern* 06/2024 – 08/2024

- Investigated and configured the Promtail-Loki-Grafana stack for microservices on **AWS ECS** clusters, resulting in a monthly savings of **\$36K** on CloudWatch logs
- Deployed **Promtail** logging agents on **EC2** instances to efficiently collect application logs
- Implemented a **distributed Loki** setup to independently scale read and write traffic

**TikTok** San Jose, USA  
*Software Engineer Intern* 06/2022 – 08/2022

- Developed a streaming data pipeline using **Flink**, handling a peak daily throughput of 450,000 transactions for global ad auditing
- Created a **Lambda** architecture to integrate offline batch processing with online streaming events for real-time query support
- Built an extensible pipeline utilizing message queues like **Kafka** and **RocketMQ**, and OLAP databases such as **ClickHouse** and **Hive**

**Nokia** Hangzhou, China  
*Software Engineering Intern* 01/2021 – 03/2021

- Developed an internal tool for searching and displaying errors in 5G technical specifications, assisting over **1,000** engineers globally with **18,000+** corrections.
- Implemented a CI/CD pipeline and release strategy on development and production servers with **Jenkins** and **GitLab**
- Designed a systematic testing framework using **PyTest**, incorporating GitLab Flow for unit testing and static code analysis.

## PROJECTS

---

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

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

## RESEARCH

---

**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”, Proceedings of the VLDB Endowment, 17(5), 1146-1158 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 online classes for underprivileged kids in Dandong 11/2017 – 06/2019