

# Pengyu Wang

Website: phillipw.info

Github: github.com/cswpy

Email: pengyu.wang@yale.edu

LinkedIn: pengyu-wang

Available from May 15 to August 30, 2024

## EDUCATION

### Yale University

*Master of Science in Computer Science (thesis track)*

New Haven, USA

09/2023 – 05/2025

### New York University

*Bachelor of Science – Computer Science; Cumulative GPA: 3.89, Major GPA: 3.95*

Abu Dhabi, UAE

09/2019 – 05/2023

## EXPERIENCE

### TikTok

*Software Engineering Intern*

San Jose, USA

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

*Software Engineering Intern*

Hangzhou, China

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 style checking

### combyne

*Junior Data Scientist*

Munich, Germany

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

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

**Screen Helper – a webapp that tracks pupils for better eye habits**: Developed the webcam utilization, video processing, pupil identification with Deep Learning model, and the backend. Tech: Python, OpenCV, Flask

## RESEARCH

### Improving classification performance on Whole-Slide Images

*Course Project*

Abu Dhabi, UAE

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

*Cranmore Fellowship recipient*

Abu Dhabi, UAE

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

*Capstone project*

Abu Dhabi, UAE

02/2022 – Present

- Extending relational databases to evaluate Package Queries that is a **NP-hard** problem
- Utilizing **K-Dimensional Tree** to partition **10–100** million tuples into groups within seconds
- Converting Package Queries into **Integer Linear Programming** problems that could be solved by Gurobi optimizer

## HONORS AND AWARDS

Ranked 447/10724 globally in Google Hash Code 2020

2nd Place, NYUAD International Hackathon in Quantum Computing 2022

1st Prize, Artificial Intelligence Competition of China 2018

## COMMUNITY SERVICE

### Regional Manager & Head of IT, Education Without Barriers

*Led 18+ volunteers to teach weekly classes for underprivileged kids in Dandong*

China

11/2017 – 06/2019