

Ching Pui WAN

67109110 | cpwan@connect.ust.hk | github.com/cpwan | linkedin.com/in/cpwan/

Profile Summary

A computer science professional with an MPhil in Computer Science and Engineering from HKUST and 3 years of industry experience in delivering AI-based solutions. Skilled in AI/ML, including video analytics and LLM, with a proven track record of developing and implementing innovative AI-based solutions that contribute to business growth.

Education

Hong Kong University of Science and Technology

Hong Kong

MPhil in Computer Science and Engineering – GPA:4.0/4.3

Sep 2019 - Aug 2021

- Best paper award in FTL-IJCAI'21
- Thesis: *Robust Federated Learning with Attack-Adaptive Aggregation* (Supervisor: Prof. Qifeng Chen)
- **Courses:** Deep Learning, Data mining, Optimization, Software analysis, Cryptography

Hong Kong University of Science and Technology

Hong Kong

B.Sc in Mathematics and Computer Science – GPA:3.9/4.3, First Class Honors

Sep 2015 - Aug 2019

- **Courses:** Algorithms, Software Engineering, OOP, Probability, Statistical Inference, Analysis, Algebra

ETH Zürich

Switzerland

Exchange Program – GPA:5.3/6.0

Sep 2018 - Jan 2019

- **Courses:** Deep Learning, Advanced Machine Learning, Computer Vision

Work Experience

Lenovo PCCW Solutions

Hong Kong

Data Engineer

Jun 2023 - Present

- Lead projects on Large Language Models (LLM) and video analytics.
- Built private LLM application prototypes, including RAG question answering engine, recommendation engine, database querying agents on enterprise data for aviation and banking industry, with LangChain and Huggingface.
- Deployed private LLM engine on Docker container on Ubuntu on cloud/ bare-metal/ VMWare/ Proxmox, scaling across GPU clusters with vLLM and Ray, and benchmarked inference speed for Llama, Mistral models.
- Designed, fine-tuned, and deployed object detection pipelines on edge computing devices (Nvidia Jetson) and CCTV cameras for video analytics applications in aviation industry, including human intrusion detection, trolley detection, refuse room management.

Lenovo Machine Intelligence Center

Hong Kong

Decision Intelligence Researcher

Sep 2021 - May 2023

- Developed a lightweight document layout parsing tool using the YOLOX object detection model, achieving a 5 times speedup and being 10 times smaller than state-of-the-art methods.
- Developed a table extraction model with a visual transformer by reformulating the task as an image-to-sequence problem. Achieved robust performance on diverse styles of tables while maintaining fast processing speed.
- Delivered a PoC to industry partners on engineering diagram recognition.
- Developed neural combinatorial optimization models using deep reinforcement learning and training speed-up techniques. Reduced solving time from hours to seconds compared to traditional solvers, while maintaining performance with only 3 hours of training.
- Coached 4 interns to achieve the 3rd place in the NeurIPS 2022 Vehicle Routing Competition at the time of their internship completion.

Hong Kong University of Science and Technology

Hong Kong

Teaching Assistant

Feb 2020 - Dec 2020

- Delivered C++ tutorials for the Object Oriented Programming course.
- Designed assignments and held a contest for the Advanced Machine Learning course.

Google LA x Institute for Pure and Applied Mathematics

Los Angeles

Research Intern

Jun 2018 - Aug 2018

- Project on "Extreme rare events classification with predicted positive data".
- Developed algorithms for effective ensemble classifiers on limited data, tuning the classifiers with only predicted positive samples.
- Predicted the optimal thresholds of the classifiers to improve advertising security.

Microsoft Research Asia x HKUST

Research Intern

Hong Kong

Jun 2017 - Aug 2017

- Project on “Efficient Communication in Distributed Machine Learning”.
- Implemented distributed machine learning algorithms using Microsoft’s CNTK framework on C++.
- Conducted experiments on image classification, recommendation systems, and language translation tasks.

Skills

Programming	Python, C++, Java
Video Analytics	PyTorch, OpenCV, MMDetection, PaddlePaddle, ONNX, CVAT
Gen AI	LangChain, LlamaIndex, Azure OpenAI, Google Vertex AI
Data Science	Pandas, scikit-learn, Plotly, Gradio, Streamlit
Software Engineering	Docker, Git, Azure DevOps, Github action, Gitlab CI/CD
Web	FastAPI, Flask

Achievements

2021 **Best Paper Award**, FTL-IJCAI’21

Montreal

2017 **Best Entrepreneur Award**, HKUST x RADICA Big Datathon

Hong Kong

Publications

JOURNAL ARTICLES

RLOR: A Flexible Framework of Deep Reinforcement Learning for Operation Research

Ching Pui Wan, Tung Li, Jason Min Wang

arXiv:2303.13117 (2023)

Robust Federated Learning with Attack-Adaptive Aggregation

Ching Pui Wan, Qifeng Chen

FTL-IJCAI’21 (2021)

Languages

English	Professional proficiency
Cantonese	Native proficiency
Mandarin	Professional proficiency