

KUEI-CHUN KAO

• [linkedin.com/in/kuei-chun-cao](https://www.linkedin.com/in/kuei-chun-cao) • [Github](#) • [Personal Website](#) • (310) 621-1492 • johnson0213@ucla.edu

EDUCATION

University of California, Los Angeles (UCLA)

M.S. in Computer Science, Specialization in Artificial Intelligence
Graduate Student Researcher advised by *Prof. Cho-Jui Hsieh*

Los Angeles, CA
Sept. 2023 – Present

National Yang-Ming -- Chiao-Tung University (NYCU)

B.S. in Computer Science, Overall GPA: 3.99/4.3 (3.84/4.0)
Scholarship for Academic Excellence performance 2 times (1% of computer science department per semester)
Teaching Experience: *Teaching Assistant of Introduce to Natural Language Preprocessing (2022 Spring)*

Hsinchu, Taiwan
Sept. 2018 – Dec. 2022

WORK EXPERIENCE

Appier

Machine Learning Scientist (AI Real Team Bidding Team)

Taipei, Taiwan
Sept. 2021 – July 2023

- Improved User Lookalike models to produce the distinguished user score for each unique user id based on existing client site activities and deployed models online by using CI/CD pipelines, enhancing **20%** improvement on AUROC compared to baseline.
- Extracted user behavior patterns from **1000K+** conversion funnel data and analyzed the Click-Through Rate of different campaigns and supply side platforms using PySpark, SQL and Pandas, resulting in **120%** revenue and Cost per Click growth within 3 months.
- Surveyed and used calibration and auto-tuning approaches on our bidding models, increasing model stability on unseen impression data.
- Implemented a dashboard via Grafana for monitoring and alerting outlier training data, saving up **30%** of the model trouble shooting time.
- Integrated and refactored different data serving pipelines via Spark and Kubernetes, causing **40%** of data pipeline storage reduction.

Umbo CV

AI Engineer Intern

Taipei, Taiwan
Aug. 2021 – Sept. 2021

- Investigated research and employed model compression techniques such as structural pruning and quantization to person re-identification model on real-time streaming cameras, speeding up **20%** of FPS and **3x times** fewer FLOPs, while keeping the false negative the same.

Cinnamon AI

AI Bootcamp Summer Intern

Taipei, Taiwan
July 2021 – Aug. 2021

- Cooperated with 3 interns to work on an AI-based trip advisor project, which constructed an ML pipeline to enable deep learning models and built a database system to store user profile and tourist attraction, reducing **50%** of ML engineers' deployment time.
- Implemented a Seq2Seq-based model to recommend tourist attractions based on customer's requirement and arrange suitable trip routes deployed on Gradio.io to make a fast user interface and model serving.

SELECTED PROJECT

Algebra Word Problem Solver

July 2021 – Mar. 2023

- Proposed a new structural deep learning model to solve Algebra Word Problem Solver and generated the intermediate form of equation index, achieving 10% improvement in equation accuracy and problem accuracy on both English and Chinese algebraic datasets.
- Designed a new training objective and conducted experiments, which adopts the concepts of the human solving strategies, generating multiple operation trees explicitly and representing the reasonable solving process behind the model's solution equation.

Predicting Smartphone Users' Kill Time Moments

July 2020 – Feb. 2023

- Leveraged deep learning fusion model to investigate users' kill time behavior based on **1000K+** mobile phone-sensor and screenshot data, which is collected by our developed Android App and uploaded to Firebase.
- Employed a two-stage clustering approach to separate users into four groups according to the patterns of their phone-usage behaviors, and then built a fusion model for each group, yielding overall strong performance on AUROC.

Model Compression for Object Detection

July 2020 – June 2021

- Applied structural pruning, knowledge distillation and quantization on YOLOv4. The developed models not only fit for embedded systems (Ex: NVIDIA Jetson TX2) but also achieve higher FPS and mAP at the same time on the multi-spectral infrared dataset.
- Winner of the award: "2021 ACM ICMR Embedded Deep Learning Object Detection Model Compression Competition for Traffic in Asian Countries" -- **Final Round (5th place)**.

SKILLS

- Programming:** C/C++, Python (Package: PyTorch, Tensorflow, PySpark, PyTest), SQL, Shell Script, Scala (Spark), MATLAB, R
- DevOps & Tools:** GCP, Docker, K8s, Git, Jenkins, Argo CD, Helm, Prometheus, Airflow, Grafana, Clickhouse, InfluxDB, MySQL

PUBLICATIONS

- Yu-Chun Chen*, **Kuei-Chun Kao***, Yu-Jen Lee*, Yung-Ju Chang "Does Receiving Less Personally Relevant but Attention-demanding Notifications while 'Killing Time' Increase Engagement? An Exploratory Study", **CHI'24** (Under Review, * indicates equal contribution)
- Kuei-Chun Kao**, Chao-Chun Liang, Keh-Yih Su "Knowledge-Guided Algebra Word Problem Solving", **EMNLP'24** (Revise & Resubmit)
- Yu-Chun Chen, Yu-Jen Lee, **Kuei-Chun Kao**, Jie Tsai, En-Chi Liang, Wei-Chen Chiu, Faye Shih, Yung-Ju Chang "[Are You Killing Time? Predicting Smartphone Users' Time-killing Moments via Fusion of Smartphone Sensor Data and Screenshots](#)", **CHI'23**
- Yu-Chun Chen*, **Kuei-Chun Kao***, Yu-Jen Lee, Faye Shih, Wei-Chen Chiu, Yung-Ju Chang "[Killing-Time Detection from Smartphone Screenshots](#)", **UbiComp'21** (* indicates equal contribution)