# TAICHANG ZHOU

(424) 407 5317◇ tzhouam@ucla.edu

## EDUCATION

**University of California, Los Angeles** Sep 2023 - May 2025 (expected)
*M.S. in Computer Science*

**Hong Kong University of Science and Technology** Sep 2019 - May 2023
*B.Eng. in Computer Science and Engineering* *Overall GPA: 3.7/4.3(top 10%)*

**Georgia Institute of Technology** Jan 2022 - May 2022
*Exchange* *Overall GPA: 4.0/4.0*

## SKILL-SET

**Programming Language:** C/C++, Python, Java, HTML, CSS, JavaScript(JS), SQL

**Machine Learning:** Numpy, Pandas, Pytorch, CNN, Transformer, LSTM, Large Language Model(LLM)

**Other:** Cloud Computing(HDFS,Hadoop,Spark), Recommendation System, Blockchain, Computer Network

## INTERNSHIP

**Vetrackr** Jun 2022 - Aug 2022
*Junior AI Developer* *Hong Kong*

- Collaborated with the team to contribute to risk management model of predicting taxi drivers' security ranks
- Retrieved the weather data through Python web crawler, and collected driving data through plug-in detectors on taxis
- Implemented algorithms based on GPS to identify distinctive driving behaviors with over 90% accuracy.
- Established a deep learning system from scratch with multi-layer LSTM to predict driver risk appetite and recommended speed based on weather conditions, attaining a precision over 40%
- Evaluated the risk prediction model with saliency analysis to ensure the reliability and robustness

## PROJECT

**Extremely Efficient Fine-Tuning for Shallow Alignment Using One Token** Oct 2023 - June 2024
*Co-Author* *Submitted to ACL ARR*

- Participate in the development of a novel fine-tuning technique using only one single global token, achieving comparable performance to common PEFT methods like LoRA.
- Conducted extensive experiments on Llama models, analyzing the harmless-helpless trade-off using the Beaver dataset.
- Demonstrated over 95% reduction in trainable parameters compared to Lora fine-tuning while maintaining comparable performance.
- Contributed to potential breakthroughs in efficient model adaptation, relevant to large-scale AI applications in social media platforms.

**Glancee - Adaptive Learning System** Sep 2020 - Jul 2021
*Co-Author* *Accepted by CHI '22: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*

- Constructed an adaptive learning system for instructors to grasp student learning status in synchronous online classes
- Performed crapping data, pre-processed the data including data cleaning and data aggregation, and labeled the dataset of emotion-expression pair videos
- Assisted in implementation of Openface to identify 68 face landmarks, and emotion/engagement detectors to identify students' emotions from facial landmarks
- Implemented front-end page for 3 views of the system by HTML, JS, and CSS: dashboard view (outlook adjustment), in-class view (status collection from student and illustration for lecturer) and post-class view (review for lecturer)

**GLORIA: Generative World Model with Latent Object Representations and Actions** Otc 2023 - Present

- Developed a novel generative world model that uses latent object representations and actions to learn from video data without explicit agent-environment distinctions.
- Implemented an object-centric approach, treating each object as an agent to enhance causal imagination, planning, and physical alignment.
- Reformulated the world model to focus on object-level actions and dynamics, enabling learning from diverse, unannotated video data.
- Achieved improved computational efficiency and scalability by reducing token usage and aligning with true physical dynamics.
- Demonstrated superior generation quality, diversity, and planning abilities through qualitative and quantitative experiments across various datasets.
- Proposed future work directions, including tuning object detection models and extending the model to 3D generation and text conditioning.