Tzu-Heng (Brian) Huang

Email: **thuang273@wisc.edu**
Seeking for 2024 Research Internship

LinkedIn: **zihengh1**
Phone Number: +1 608-960-6927
Personal Webpage: **zihengh1.github.io/**

Education

- **University of Wisconsin-Madison** — Madison, Wisconsin
  3rd-year Ph.D. student in Computer Science. Minoring in Economics. — *Aug. 2021 — Jun. 2026 (expected)*
- **National Chengchi University (NCCU)** — Taipei, Taiwan
  B.S. in Computer Science. (Major GPA: 3.96/4.30) — *Sep. 2016 — Jul. 2020*

Research Summary

I am passionate about advancing potent and efficient techniques for ***data-centric AI***, specifically empowering ***foundation models*** to acquire more knowledge while diminishing reliance on extensive human annotations. To achieve this objective, I have undertaken several initiatives:

1. Pioneered the first ***parameter trading market*** to enhance multi-agents' performance while reducing training costs [1].
2. Proposed a scalable ***multimodal data curation framework*** via ensembles and object detection to improve CLIP [3].
3. Built a ***zero-cost labeling system***, integrating with synthesized labeling sources via ICL, RAG, and Code LLMs [4].
4. Presented a new benchmarking framework for ***automated weak supervision techniques*** on diverse science tasks [5].
5. Tailored LLaMA for low-resource domains and built the first ***large language model*** for traditional Chinese medicine.

Publications

[1] Tzu-Heng Huang, Harit Vishwakarma, Frederic Sala, "Train 'n Trade: Foundations of Parameter Markets", *in Neural Information Processing Systems (NeurIPS), 2023.*

[2] Nicholas Roberts, Xintong Li, Dyah Adila, Sonia Cromp, **Tzu-Heng Huang**, Jitian Zhao, Frederic Sala, "Geometry-Aware Adaptation for Pretrained Models", *in Neural Information Processing Systems (NeurIPS), 2023.*

[3] Tzu-Heng Huang*, Changho Shin*, Sui Jiet Tay, Dyah Adila, Frederic Sala, "Multimodal Data Curation via Object Detection and Filter Ensembles", *in ICCV Workshop: Towards the Next Generation of Computer Vision Datasets, 2023.*

[4] Tzu-Heng Huang, Catherine Cao, Spencer Schoenberg, Harit Vishwakarma, Nicholas Roberts, Frederic Sala, "ScriptoriumWS: A Code Generation Assistant for Weak Supervision", *in ICLR Workshop: Deep Learning For Code, 2023* and *in Midwest Machine Learning Symposium, 2023.*

[5] Nicholas Roberts, Xintong Li, **Tzu-Heng Huang**, Dyah Adila, Spencer Schoenberg, Cheng-Yu Liu, Lauren Pick, Haotian Ma, Aws Albarghouthi, Frederic Sala, "AutoWS-Bench-101: Benchmarking Automated Weak Supervision with 100 Labels", *in Neural Information Processing Systems (NeurIPS), 2022.*

[6] Tzu-Heng Huang, Cheng-Hsien Tsai, Man-Kwan Shan, "Key Sensor Discovery for Quality Audit of Air Sensor Networks", *in ACM International Conference on Mobile Systems, Applications, and Services (MobiSys), 2020.*

Research Experience

- **Department of Computer Science, UW-Madison** — Madison, Wisconsin
  Graduate Research Student, advised by Prof. Frederic Sala — *Feb. 2022 — Present*
  - **Train 'n Trade: Foundations of Parameter Markets [1]** :
    - Designed a new knowledge exchange marketplace for multi-agents to trade parameter sets and to accelerate training process.
    - Applied parameter alignment to merge models and proved efficacy via faster convergence and successful monetization.
  - **Geometry-Aware Adaptation for Pretrained Models [2]**:
    - Proposed a new adaption method by leveraging relational information in label spaces to improve foundation models (CLIP).
    - Validated enhanced foundation model over SimCLR on ImageNet and scaled to extreme classes without additional training.
  - **Multimodal Data Curation via Object Detection and Filter Ensembles [3]**:
    - Built a scalable multimodal data curation framework via filter ensembles and object detection (e.g. DINO) to improve CLIP.
    - **Rank #1** on the small-scale filtering track of ICCV'23 Datacomp leaderboard.
  - **ScriptoriumWS: A Code Generation Assistant for Weak Supervision [4]**:
    - Applied in-context learning, RAG and Code LLMs to develop zero-cost labeling systems with synthesized labeling sources.
    - Showcased framework efficacy by comparable label correctness but higher coverage on par with human annotation standards.
  - **AutoWS-Bench-101: Benchmarking Automated Weak Supervision with 100 Labels [5]**:
    - Developed a new benchmarking framework to evaluate automated weak supervision techniques on diverse science tasks.

- **Awan.AI (Startup)** — San Jose, California
  CEO and Founder, collaborated with TechTCM — *May. 2023 — Present*
  - **Large Language Models for Traditional Chinese Medicine**:
    - Tailored generative AI solutions with LLaMA and LangChain and built the first LLM for traditional Chinese medicine.
    - Curated customized and synthesized datasets for LLM continue pretraining and fine-tuning with LoRA and Flash Attention.
  - **Vision-Language Model for Tongue Diagnosis in Traditional Chinese Medicine (in-progress)**:
    - Research on LLaVA with crowdsourcing tongue images and medical diagnosis to answer extreme multi-label syndromes.

- **Argonne National Laboratory** — Lemont, Illinois
  Research Intern, advised by Dr. Charles Catlett — *Jun. 2019 — Sep. 2019*
  - **Radiative Error Reduction for Low-cost Temperature Sensors**:
    - Researched pattern identification on time series and ensemble learning to improve calibration model performance by **25%**.

- **Department of Computer Science, NCCU** — Taipei, Taiwan
  Research Assistant at Data Mining Lab, advised by Prof. Man-Kwan Shan — *Sep. 2018 — Aug. 2021*
  - **Key Sensor Discovery for Quality Audit of Air Sensor Networks [6]**:
    - Proposed a new quality audit framework to monitor low-cost sensor performance while reducing human effort on inspections.
    - Developed sensor correlation model and built diverse relational graphs to discover key sensors via graph traversal algorithms.
  - **Early Prediction of Affected Sensors by Local Events Detected over Social Media**:
    - Applied spatial-temporal GNNs to detect anomalies in multivariate time series and to label diffusion among regions.
    - Developed an early prediction framework with Seq2seq models for affected region prediction with F1-score of **80%**.
  - **Missing Value Estimation of Large Scale Air Monitoring Sensor Network**:
    - Developed spatial-temporal correlation models via geo-context for missing value imputation with error rate less than **10%.**
    - Enhanced correlation models with diverse time series segmentation methodologies by **17%**.

- **Institute of Information Science, Academia Sinica** — Taipei, Taiwan
  Research Intern at Network Research Lab, advised by Dr. Ling-Jyh Chen — *Feb. 2018 — Jul. 2020*
  - **Calibrating Low-cost PM2.5 Sensors in Large Scale IoT Environmental Monitoring Systems**:
    - Proposed an adaptive calibration framework with regression models to ensure data quality of large-scale low-cost air sensors.
    - Project was awarded Undergrad Student Research Scholarship granted by Ministry of Science and Technology in Taiwan.
  - **PiM25 — Environmental Sensing Hub**:
    - Designed maker-based sensor hub with OTA updates to detect environmental conditions, e.g., air, temperature, and sounds.
    - PiM25 was accepted by HKoscon'19 and COSCUP'19 to demonstrate and was the first TW's project reported by Magpi.

Awards

- **Conference Scholar Award**: granted by NeurIPS'23.
- **ICCV'23 Datacomp Competition**: rank #1 in the small-scale filtering track.
- **First-year Departmental Scholarship**: granted by Department of Computer Science, UW-Madison.
- **International Research Intern Scholarship**: granted by National Chengchi University (NCCU).
- **Undergrad Student Research Scholarship**: granted by Ministry of Science and Technology (MOST), Taiwan.

Invited Talks

- **IoT Instantiation: Air Sensor Deployment**: invited by Nangang High School (Taipei), Dec. 2019.
- **Internship Abroad Scholarship Sharing**: invited by National Chengchi Univerisity, Sep. 2019.
- **LASS Conference International Session**: invited by Institute of Information Science, Academia Sinica, Jul. 2019.
- **Techbang Magazine Sharing: PiM25 Project**: invited by Techbang Magazine, Mar. 2019.
- **Raspberry Pi Jam: PiM25 Project**: invited by Raspberry Pi Foundation (Taiwan), Mar. 2019.
- **The 24th of Raspberry Pi Meetup: PiM25 Project**: invited by Raspberry Pi Foundation (Taiwan), Jan. 2019.

Academic Services

- **Co-organizer**: AutoML Cup in AutoML'23.
- **Paper Reviewer**: GLOBECOM'20, NeurIPS'23, ICLR'24, CoLLAs'24, ICML'24, DMLR.
- **Student Association of Taiwna (SAT), UW-Madison**: President, Jun. 2022 — May. 2023.
- **Student Association of Taiwan (SAT), UW-Madison**: Vice President, Jun. 2021 — May. 2022.

Programming Skills

- **Programming Languages**: Python, R, C++, SQL, LaTeX, and Shell Programming.
- **Technologies**: PyTorch, Tensorflow, Keras, ShinyApp, PostgreSQL, Linux, Flask, Dash Visualization, Git, and Vim.