Tzu-Heng (Brian) Huang

Email: **thuang273@wisc.edu**

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

Education

- **University of Wisconsin-Madison** — Madison, Wisconsin
  Third-year Ph.D. student in Computer Science. Minor in Economics. — *Aug. 2021 — Present*

- **National Chengchi University (NCCU)** — Taipei, Taiwan
  B.S. in Computer Science. (Major GPA: 3.96/4.30) — *Sep. 2016 — Jul. 2020*

Research Interests

I am passionate about advancing machine learning by building effective techniques that empower models to **acquire more knowledge with reduced supervision**. For example, I have developed a knowledge transfer/trading framework involved with multi-agents to buy and sell parameter sets to enhance downstream task performance while alleviating the training cost burden [1]. Additionally, I have delved into the realm of data-centric machine learning, specializing in the design of efficient methodologies in multimodal data curation [3], synthesized label sources via prompting [4], and automated labeling benchmark [5]. These strategies are rooted in the utilization of weak supervision frameworks, enabling the construction and exploitation of foundational models (CLIP, Code LLMs, etc.) with reduction in the need for extensive human annotations.

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 viable knowledge transfer marketplace for multi-agents to trade parameter sets and to reduce training expenses.
    - Demonstrated framework effectiveness under diverse practical scenarios with faster convergence and provable bounds.

  - **Geometry-Aware Adaptation for Pretrained Models [2]**:
    - Proposed a new adaption technique by leveraging limited relational information in label spaces to improve pretrained models.

  - **Multimodal Data Curation via Object Detection and Filter Ensembles [3]**:
    - Developed a novel data curation methodology via ensembling for multimodal datasets and improved vision-language models.
    - Rank #1 on the small-scale filtering track of ICCV'23 Datacomp competition leaderboard.

  - **ScriptoriumWS: A Code Generation Assistant for Weak Supervision [4]**:
    - Proposed a low-cost data labeling framework with weak supervision by leveraging Code LLMs to synthesize label sources.

  - **AutoWS-Bench-101: Benchmarking Automated Weak Supervision with 100 Labels [5]**:
    - Developed a new benchmark to evaluate automated weak supervision techniques in diverse application domains.

- **Awan.AI (Startup)** — San Jose, California
  CEO and Co-founder, collaborated with TechTCM — *May. 2023 — Present*

  - **Large Language Models for Traditional Chinese Medicine**:
    - Customized low-cost generative AI (with LLaMa family) and built the first language model for traditional Chinese medicine.
    - Curated customized and synthesized datasets for continue pretraining and LLM fine-tuning with LoRA and Flash Attention.

  - **Vision Language Model for Tongue Diagnosis in Traditional Chinese Medicine**:
    - Research on CLIP model with crowd-sourcing tongue images and medical diagnosis to detect 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 novel quality audit framework to inspect sensor performance and to reduce the cost of human inspections.
    - Developed sensor correlation models and built relational graphs to discover key sensors via approximation algorithms.
  - ○ **Early Prediction of Affected Sensors by Local Events Detected over Social Media**:
    - Leveraged spatial-temporal GNN models to detect anomalies in multivariate time series and to label affected timestamps.
    - Developed an early prediction framework with BiGRU/BiLSTM 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 for missing value imputation with error rate less than **10%.**
    - Improved correlation models through sequential clustering in time series segmentation 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 sensors.
    - Project was awarded Undergrad Student Research Scholarship granted by the Ministry of Science and Technology, Taiwan.
  - ○ <u>**PiM25**</u> **– Environmental Sensing Hub**:
    - Designed a maker-based sensor hub with over-the-air updates to detect various environmental conditions.
    - Deployed on-device pretrained audio models to recognize environmental sounds with F1-score of **75%**.
    - PiM25 was accepted by **HKoscon'19** and **COSCUP'19** to demonstrate and was the first TW's project reported by <u>Magpi</u>.

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 the 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.
- **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.