



Tzu-Heng (Brian) Huang

✉ thuag273@wisc.edu

in [zihengh1](#)

🌐 zihengh1.github.io

Education

- 2021 – 2026 (Expected)  **Ph.D. in Computer Science. University of Wisconsin-Madison.**
Advised by Prof. Frederic Sala. Minor in Economics.
- 2016 – 2020  **B.S. in Computer Science. National Chengchi University, Taiwan.**
Advised by Prof. Man-Kwan Shan and Dr. Ling-Jyh Chen. Major GPA: 3.96




Research Interests






My research centers on *data-centric AI for multimodal foundation models*, to enable models to learn more with less supervision. Several works I have developed, including (1) model-aware data selection for efficient pretraining (e.g., training CLIP on hundreds million data), (2) data curation for multimodal models through objective detection (*1st place on the Datacomp leaderboard*), (3) a 500x cheaper auto-labeling system over LLM annotators (*NeurIPS'24 Spotlight*), and (4) a new accelerated training framework via model merging.

Research Experience

- May. 2024 – Dec. 2024  **AIML Research Intern.** Apple.
advised by Dr. Javier Movellan and Manjot Bilkhu.
— *Automated Model-aware Data Selection for Efficient Pretraining.*
- Aug. 2021 – Present  **Graduate Research Student.** UW-Madison.
advised by Prof. Frederic Sala.
— *Data-centric AI for Foundation Models: Auto-labeling and Data Curation.*
— *Parameter Marketplace: Through Model Merging and Auction Agents.*
- May. 2023 – Apr. 2024  **Founder & CEO.** Awan.AI LLC.
collaborated with and funded by TechTCM.
— *LLM for Traditional Chinese Medicine and Tongue Syndrome Diagnosis.*
- Jun. 2019 – Sep. 2019  **Research Intern.** Argonne National Laboratory.
advised by Dr. Charlie Catlett and Dr. Rajesh Sankaran.
— *Ensemble-based Time Series Calibration for Low-cost Sensors.*
- Sep. 2018 – Aug. 2021  **Research Assistant.** National Chengchi University.
advised by Prof. Man-Kwan Shan.
— *Spatio-temporal Modeling in Large-scale Sensor Networks.*
- Feb. 2018 – Jul. 2020  **Research Intern.** Academia Sinica.
advised by Dr. Ling-Jyh Chen.
— *Large-scale Air Quality Sensor Network Development.*






Research Publications

- 1 T.-H. Huang, C. Cao, V. Bhargava, and F. Sala, “The ALCHEmist: Automated Labeling 500x CHEaper than LLM Data Annotators,” in *Neural Information Processing Systems (NeurIPS)* [*Spotlight*], 2024.
 URL: <https://arxiv.org/abs/2407.11004>.
- 2 W. Tan, N. Roberts, T.-H. Huang, *et al.*, “MoRe Fine-Tuning with 10x Fewer Parameters,” in *ICML Workshop: Efficient Systems for Foundation Models (ES-FoMo) and ICML Workshop: Foundation Models in the Wild.*, 2024.  URL: <https://arxiv.org/abs/2408.17383>.
- 3 T.-H. Huang, C. Cao, S. Schoenberg, H. Vishwakarma, N. Roberts, and F. Sala, “ScriptoriumWS: A Code Generation Assistant for Weak Supervision,” in *ICLR Workshop: Deep Learning For Code (DL4C)*, 2023.
 URL: <https://dl4c.github.io/assets/pdf/papers/30.pdf>.







- 4 T.-H. Huang, C. Shin, S. J. Tay, D. Adila, and F. Sala, "Multimodal Data Curation via Object Detection and Filter Ensembles," in *ICCV Workshop: Towards the Next Generation of Computer Vision Datasets (TNGCV)* [**1st place on the Datacomp leaderboard (small-scale filtering track)**], 2023.  URL: <https://arxiv.org/abs/2401.12225>.
- 5 T.-H. Huang, H. Vishwakarma, and F. Sala, "Train 'n Trade: Foundations of Parameter Markets," in *Neural Information Processing Systems (NeurIPS)*, 2023.  URL: <https://arxiv.org/abs/2312.04740>.
- 6 N. Roberts, X. Li, D. Adila, *et al.*, "Geometry-Aware Adaptation for Pretrained Models," in *Neural Information Processing Systems (NeurIPS)*, 2023.  URL: <https://arxiv.org/abs/2307.12226>.
- 7 N. Roberts, X. Li, T.-H. Huang, *et al.*, "AutoWS-Bench-101: Benchmarking Automated Weak Supervision with 100 Labels," in *Neural Information Processing Systems (NeurIPS)*, 2022.  URL: <https://arxiv.org/abs/2208.14362>.
- 8 T.-H. Huang, C.-H. Tsai, and M.-K. Shan, "Key Sensor Discovery for Quality Audit of Air Sensor Networks," in *ACM International Conference on Mobile Systems, Applications, and Services (MobiSys)*, 2020.  URL: <https://dl.acm.org/doi/abs/10.1145/3386901.3396606>.

Miscellaneous





Awards and Achievements

- | | |
|------|---|
| 2023 |  ICCV Datacomp Competition , won the first place in the small-scale filtering track.
 Scholar Award , granted by NeurIPS'23. |
| 2021 |  First-year Departmental Scholarship , granted by UW-Madison. |
| 2020 |  Research Intern Scholarship , granted by National Chengchi University.
 Undergrad Research Scholarship , granted by Ministry of Science and Technology. |



Invited Talks

- | | |
|-----------|--|
| Dec. 2019 |  IoT Project Development , invited by Nangang High School (Taiwan). |
| Sep. 2019 |  Intern Research Talk , invited by National Chengchi University. |
| Jul. 2019 |  LASS Conference: International Session , invited by Academia Sinica. |
| Mar. 2019 |  Techbang Magazine: PiM25 Project , invited by Techbang Magazine.
 Raspberry Pi Jam: PiM25 Project , invited by Raspberry Pi Foundation (Taiwan). |
| Jan. 2019 |  Raspberry Pi Meetup: PiM25 Project , invited by Raspberry Pi Foundation (Taiwan). |

Academic Services

- | | |
|----------------|--|
| 2023 – Present |  Paper Reviewer . NeurIPS'23 & 24, ICLR'24 & 25, CoLLAs'24, ICML'24, DMLR. |
| 2023 |  Co-organizer . AutoML Cup in AutoML Conference. |
| 2022 – 2023 |  President of Student Association of Taiwan , UW-Madison. |
| 2021 – 2022 |  Vice President of Student Association of Taiwan , UW-Madison. |

Skills

- | | |
|-----------------------|--|
| Programming Languages |  Python, R, C++/C, SQL, \LaTeX , and Shell Programming. |
| Technologies |  (Distributed) PyTorch, Tensorflow, Keras, ShinyApp, PostgreSQL, and Vim. |