Changho Shin

cs1095@princeton.edu
https://ch-shin.github.io
Peretsman Scully Hall, Room 120
60 Olden St, Princeton, NJ 08540

ACADEMIC POSITIONS

Princeton University, Princeton, NJ — Feb. 2026 – Present
Postdoctoral Researcher
• Advisor: Brenden Lake

RESEARCH INTERESTS

My research focuses on **data-centric AI for foundation models**, including large language models (LLMs) and multimodal foundation models (MLLMs). I develop methods for **efficient supervision**, leveraging **weak supervision**, **data selection**, and **weak-to-strong generalization**. Additionally, I explore **inference-time steering**, such as **representation editing**, to steer foundation models at inference time, enabling robust adaptation and the adoption of new capabilities.

University of Wisconsin–Madison — Sep. 2020 – Dec. 2025
• Ph.D. in Computer Science; M.S. in Mathematics
• Advisor: Frederic Sala
• Thesis: *Learning from Weak Signals: Data-Centric Methods for Foundation Models*

Seoul National University — Mar. 2015 – Feb. 2017
• M.S. Machine Learning
• Advisor: Professor Wonjong Rhee

Seoul National University — Mar. 2011 – Feb. 2015
• B.A. in Psychology, B.S. in Computer Science and Engineering
• Graduated with honors (Cum Laude)

HONORS & AWARDS

Qualcomm Innovation Fellowship Finalist 2024
Best Paper Award Honorable Mention, NeurIPS R0-FoMo Workshop 2023
NeurIPS Scholar Award 2023
1st Place, DataComp Competition (Small Track, Filtering) 2023
CS Departmental Scholarship, University of Wisconsin-Madison 2020

CONFERENCE PUBLICATIONS

[C8] Jitian Zhao*, **Changho Shin***, Tzu-Heng Huang, Srinath Namburi, Frederic Sala, "CARE: Confounder-Aware Aggregation for Reliable LLM Evaluation", *International Conference on Machine Learning (ICML)*; 2026.

[C7] **Changho Shin**, John Cooper, Frederic Sala, "Weak-to-Strong Generalization Through the Data-Centric Lens", *International Conference on Learning Representations (ICLR)*, 2025.

[C6] Yijing Zhang, Dyah Adila, **Changho Shin**, Frederic Sala, "Personalize Your LLM: Fake it then Align it", *North American Chapter of the Association for Computational Linguistics (NAACL) Findings*, 2025.

[C5] **Changho Shin**, Jitian Zhao, Sonia Cromp, Harit Vishwakarma, Frederic Sala, "OTTER: Improving Zero-Shot Classification via Optimal Transport", *Neural Information Processing Systems (NeurIPS)*, 2024.

[C4] Dyah Adila*, **Changho Shin***, Linrong Cai, Frederic Sala, "Zero-Shot Robustification of Zero-Shot Models With Auxiliary Foundation Models", *International Conference on Learning Representations (ICLR)*, 2024.

Best Paper Award Honorable Mention, Oral Presentation at *NeurIPS 2023 R0-FoMo Workshop*.

[C3] **Changho Shin**, Sonia Cromp, Dyah Adila, Frederic Sala, "Mitigating Source Bias for Fairer Weak Supervision", *Neural Information Processing Systems (NeurIPS)*, 2023.

[C2] **Changho Shin**, Winfred Li, Harit Vishwakarma, Nicholas Roberts, Frederic Sala, "Universalizing Weak Supervision", *International Conference on Learning Representations (ICLR)*, 2022.

[C1] **Changho Shin**, Sunghwan Joo, Jaeryun Yim, Hyoseop Lee, Taesup Moon, Wonjong Rhee, "Subtask Gated Networks for Non-Intrusive Load Monitoring", *AAAI Conference on Artificial Intelligence*, 2019.

JOURNAL PUBLICATIONS

[J2] **Changho Shin**, Eunjung Lee, Jeongyun Han, Jaeryun Yim, Hyoseop Lee, Wonjong Rhee, "The ENERTALK Dataset, 15 Hz Electricity Consumption Data from 22 Houses in Korea", *Nature Scientific Data*, 2019 (Impact Factor = 5.929).

[J1] **Changho Shin**, Seungeun Rho, Hyoseop Lee, Wonjong Rhee, "Data Requirements for Applying Machine Learning to Energy Disaggregation", *Energies*, May 2019 (Impact Factor = 2.707).

PREPRINTS

[P3] **Changho Shin**, Xinya Yan, Suenggwan Jo, Sungjun Cho, Shourjo Aditya Chaudhuri, Frederic Sala, "TARDIS: Mitigating Temporal Misalignment via Representation Steering", *arxiv*, 2025.

[P2] Dyah Adila, **Changho Shin**, Yijing Zhang, Frederic Sala, "Alignment, Simplified: Steering LLMs with Self-Generated Preferences", *arxiv*, 2025.

[P1] Amanda Dsouza, Christopher Glaze, **Changho Shin**, Frederic Sala, "Evaluating Language Model Context Windows: A 'Working Memory' Test and Inference-time Correction", *arxiv*, 2024.

WORKSHOP PUBLICATIONS

[W7] **Changho Shin**, David Alvarez-Melis, "Curriculum Learning as Transport: Training Along Wasserstein Geodesics", *NeurIPS 2025 CCFM Workshop*.

[W6] Sungjun Cho, **Changho Shin**, Suenggwan Jo, Xinya Yan, Shourjo Aditya Chaudhuri, Frederic Sala, "LLM-Integrated Bayesian State Space Models for Multimodal Time-Series Forecasting", *NeurIPS 2025 BERT2S Workshop*.

[W5] Dyah Adila, **Changho Shin**, Yijing Zhang, Frederic Sala, "Is Free Self-alignment Possible?", *NeurIPS 2024 MINT Workshop*.

[W4] **Changho Shin***, Joon Suk Huh*, Elina Choi, "Pool-Search-Demonstrate: Improving Data-wrangling LLMs via better in-context examples", *NeurIPS 2023 TRL Workshop*. **Oral Presentation**.

[W3] **Changho Shin***, Tzu-heng Huang*, Sui Jiet Tay, Dyah Adila, Frederic Sala, "Multimodal Data Curation via Object Detection and Filter Ensembles", *ICCV 2023 Datacomp Workshop* (Rank #1 in DataComp competition filtering track (small)).

[W2] **Changho Shin**, Alice Schoenauer-Sebag, "Can we get smarter than majority vote? Efficient use of individual rater's labels for content moderation", *NeurIPS 2022 ENLSP Workshop*.

JOB EXPERIENCE

Microsoft Research, Cambridge — Jun. 2025 – Aug. 2025
Research Intern
- Mentor: David Alvarez-Melis
- Project: *Curriculum Learning as Transport: Training Along Wasserstein Geodesics*

Snorkel AI, California — Jun. 2024 – Aug. 2024
Research Intern
- Mentor: Christopher Glaze, Paroma Varma
- Projects: *Reward Modeling, Synthetic Data Generation, LLM Evaluation*

Twitter, San Francisco — Jun. 2022 – Aug. 2022
ML Engineer Intern
- Mentor: Alice Schoenauer Sebag

• Manager: Milind Ganjoo
• *Improving toxicity classification via weak supervision [W1]*

Encored Technologies, Seoul, Korea Jan. 2018 – Jul. 2020
Data Scientist
• Manager: Hyoseop Lee
• *Non-intrusive load monitoring [C1, J1, J2], Energy forecasting*

Korea Institute for Defense Analyses, Seoul, Korea Jan. 2017 – Dec. 2017
Researcher

TEACHING EXPERIENCE

University of Wisconsin-Madison
• Teaching assistant for CS 839 (Foundation Models) Fall 2023
• Teaching assistant for CS 300 (Programming II) Fall 2022, Spring 2023
• Teaching assistant for CS 760 (Machine Learning) Fall 2021, Spring 2022
• Teaching assistant for CS 320 (Data Programming II) Spring 2021
• Teaching assistant for CS 220 (Data Programming I) Fall 2020

TECHNICAL SKILLS

Machine Learning / Deep Learning / Data Science
PyTorch, TensorFlow, Keras, scikit-learn, NumPy, Pandas, SciPy

DBMS
MySQL, MongoDB, PySpark

Research & Development Tools
Visual Studio Code, Jupyter, PyCharm, Docker, GitHub, CircleCI, Shell, AWS

Programming Languages
Python, R, MATLAB, Java, Go, C, LaTeX