

EDUCATION	Carnegie Mellon University	Pittsburgh, United States
	<i>Ph.D. Candidate in Electrical and Computer Engineering</i>	2021 - 2025 (<i>expected</i>)
	<ul style="list-style-type: none"> • Advisor: Prof. Bhiksha Raj • Research area: Learning with Imperfect Data, Large Language Model (LLM) 	
	Carnegie Mellon University	Pittsburgh, United States
	<i>M.S. in Electrical and Computer Engineering</i>	2020 - 2021
INDUSTRIAL EXPERIENCES	<ul style="list-style-type: none"> • Advisor: Prof. Marios Savvides 	
	The University of Edinburgh	Edinburgh, United Kindom
	<i>B.Eng. in Electronics and Electrical Engineering</i>	2017 - 2019
	<ul style="list-style-type: none"> • Advisor: Prof. Sotirios Tsafaris 	
	Tianjin University	Tianjin, China
	<i>B.S. in Automation</i>	2015 - 2019
	AMD, GenAI Team Research Intern.	2024.09 - 2024.12
	<ul style="list-style-type: none"> • Continuous encoders for understanding and generation in mulit-modality models. 	
	Microsoft Resaerch, Phi Team Research Intern.	2024.05 - 2024.08
	<ul style="list-style-type: none"> • LLM-based text diversity metric and synthetic data for predicting LLM pre-training. 	
RESEARCH INTEREST	Microsoft Resaerch Asia, Social Computing Research Intern.	2023.04 - 2023.10
	<ul style="list-style-type: none"> • Framework for <i>Noisy Model Learning</i> and <i>Catastrophic Inheritance</i>. 	
	Bytedance, AI Lab Research Engineer	2021.01 - 2021.08
	<ul style="list-style-type: none"> • Large-scale fine-grained product classification and retrieval. 	
	<p>Pre-Training Data Imperfection and Physics of Foundation Models:</p> <p>Understanding and Mitigating Catastrophic Inheritance in Foundation Models:</p> <p>Leveraging Imperfect Data and Labels for Foundation Models Transferring:</p>	
AWARDS AND HONORS	• Grant of Accelerating Foundation Models Research , Microsoft	2024.01
	• Star of Tomorrow , Microsoft Research Asia	2023.10
	• Ewart Farvis Project Prize , The University of Edinburgh	2019.05
OPEN-SOURCE PROJECTS	<ul style="list-style-type: none"> • Promptbench (2.4K stars): a unified evaluation framework for LLMs and vLLMs, microsoft/PromptBench. 	
	<ul style="list-style-type: none"> • USB (1.3K stars): a unified semi-supervised learning codebase and benchmark, microsoft/Semi-supervised-learning. 	
	<ul style="list-style-type: none"> • TorchSSL (1.3K stars): semi-supervised learning algorithms, TorchSSL/TorchSSL. 	

SELECTED PUBLICATIONS	<ul style="list-style-type: none">• H. Chen, Y. Han, D. Misra, X. Li, K. Hu, D. Zou, M. Sugiyama, J. Wang, B. Raj. Slight Corruption in Pre-training Data Makes Better Diffusion Models. NeurIPS 2024, Spotlight.• H. Chen, A. Shah, J. Wang, et al., R. Singh, B. Raj. Imprecise Label Learning: A Unified Framework for Learning with Various Imprecise Label Configurations. NeurIPS 2024.• H. Chen, J. Wang, L. Feng, X. Li, Y. Wang, X. Xie, M. Sugiyama, R. Singh, B. Raj. A General Framework for Learning from Weak Supervision. ICML 2024.• H. Chen, B. Raj, X. Xie, J. Wang. On Catastrophic Inheritance of Large Foundation Models. DMLR 2024.• K. Zhu, Q. Zhao, H. Chen, J. Wang, X. Xie. Promptbench: A Unified Library for Evaluation of Large Language Models. JMLR 2024.• H. Chen, J. Wang, et al., X. Xie, M. Sugiyama, B. Raj Understanding and Mitigating the Label Noise in Pre-training on Downstream Tasks. ICLR 2024, Spotlight.• H. Chen, et al., W. Ye, J. Wang, G. Hu, M. Savvides. Conv-adapter: Exploring Parameter Efficient Transfer Learning for Convnets. CVPR Workshop 2024, Oral.• H. Chen, et al., J. Wang, B. Schiele, X. Xie, B. Raj, M. Savvides. Softmatch: Addressing the Quantity-Quality Trade-off in Semi-Supervised Learning. ICLR 2023.• Y. Wang*, H. Chen*, et al., J. Wang, M. Savvides, T. Shinozaki, B. Raj, B. Schiele, X. Xie. Freematch: Self-adaptive Thresholding for Semi-Supervised Learning. ICLR 2023.• H. Chen*, Y. Wang*, Y. Fan*, et al., B. Raj, B. Schiele, J. Wang, X. Xie, Y. Zhang. USV: A Unified Semi-Supervised Learning Benchmark for Classification. NeurIPS 2022.	
TEACHING	<ul style="list-style-type: none">• TA & Co-Instructor, Introduction to Deep Learning, CMU• TA, Pattern Recognition and Machine Learning, CMU• TA, Introduction to Deep Learning, CMU• TA, Introduction to Deep Learning, CMU	<div>2024.08 - 2024.12</div> <div>2023.08 - 2023.12</div> <div>2021.08 - 2021.12</div> <div>2021.01 - 2021.03</div>
TALKS	<div>Understanding and Mitigating the Pre-training Noise, The AI Talk</div> <div>Noisy Model Learning, ML Collective</div> <div>Understanding and Mitigating the Pre-training Noise, AI times</div> <div>Noisy Model Learning, MSRA Invited Talk</div>	
COLLABORATORS	<div>Jindong Wang, assistant professor, William & Mary</div> <div>Masashi Sugiyama, professor, RIKEN AIP and University of Tokyo</div> <div>Bernt Schiele , professor, Max Planck Institute for Informatics</div> <div>Rita Singh , professor, Carnegie Mellon University</div> <div>Yue Zhang, professor, Westlake University</div> <div>Guosheng Hu, professor, University of Bristol</div>	
ACADEMIC SERVICES	<div>Reviewers for: <i>IEEE Transactions on Pattern Analysis and Machine Intelligence</i>,</div> <div><i>ACM Transactions on Intelligent Systems and Technology</i>,</div> <div><i>International Conference on Learning Representations (ICLR)</i>,</div> <div><i>Neural Information Processing Systems (NeurIPS)</i>,</div> <div><i>Conference on Computer Vision and Pattern Recognition (CVPR)</i>,</div> <div><i>International Conference on Machine Learning (ICML)</i>,</div> <div><i>ACM KDD</i>, <i>AAAI</i></div>	