

Dang Nguyen

☎ (+1) 310-254-4895 | ✉ nguyentuanhaidang@gmail.com | 🏠 hsgser.github.io | 📺 hsgser | 🌐 dang-nguyen-50b7a7a0 | 🎓 Dang Nguyen

Research interests

My research focuses on improving data quality to enhance the performance and efficiency of large (vision-)language models. Specifically, I work on synthetic data generation and data selection to optimize training, making these models more effective and accessible.

Education

University of California, Los Angeles

California, USA

Ph.D. in Computer Science

Sep. 2023 - Present

- Advisor: Professor Baharan Mirzasoleiman
- UCLA Graduate Dean's Scholar Award

Toyo University

Tokyo, Japan

B.S. in Information Networking for Innovation and Design

Apr. 2017 - Mar. 2021

- Toyo Top Global Scholarship A
- GPA: 4.27/4.3, Top 1/300 in the faculty
- Top thesis award

Experience

Google Research

California, USA

Student Researcher

Sep. 2024 - Present

- Topic: Synthetic data generation for LLMs and Data selection for LVLMs

Cisco

California, USA

PhD Research Intern

Jun. 2024 - Sep. 2024

- Supervisor: Dr. Ali Payani
- Topic: LLM Hallucination

VinAI

Hanoi, Vietnam

AI Resident

Oct. 2020 - Aug. 2023

- Supervisor: Professor Nhat Ho (UT Austin)
- Topics: Optimal Transport and Model Merging
- Participated in an applied project which aims to improve the performance of object detectors in low-light conditions.
- Managed GPU resources for the VinAI Residency Program.

FPT Japan Holdings

Yokohama, Japan

Part-time Machine Learning Engineer

Oct. 2019 - Sep. 2020

- Participated in a long-term demand forecasting project for a chain pharmacy company in Japan.

Publications

(*) denotes equal contribution

1. **D. Nguyen***, Z. Li*, M. Bateni, V. Mirrokni, M. Razaviyayn, and B. Mirzasoleiman. Synthetic Text Generation for Training Large Language Models via Gradient Matching. *International Conference on Machine Learning (ICML)*, 2025.
2. **D. Nguyen**, W. Yang, R. Anand, Y. Yang and B. Mirzasoleiman. Mini-batch Coresets for Memory-efficient Language Model Training on Data Mixtures. *International Conference on Learning Representations (ICLR)*, 2025.
3. **D. Nguyen**, P. Haddad, E. Gan, and B. Mirzasoleiman. Changing the Training Data Distribution to Reduce Simplicity Bias Improves In-distribution Generalization. *Advances in Neural Information Processing Systems (NeurIPS)*, 2024.
4. Y. Xue, J. Siddharth, **D. Nguyen**, and B. Mirzasoleiman. Understanding the Robustness of Multi-modal Contrastive Learning to Distribution Shift. *International Conference on Learning Representations (ICLR)*, 2024.
5. K. Nguyen*, **D. Nguyen***, N. Ho. Self-Attention Amortized Distributional Projection Optimization for Sliced Wasserstein Point-Cloud Reconstruction. *International Conference on Machine Learning (ICML)*, 2023.
6. **D. Nguyen**, T. Nguyen, K. Nguyen, D. Phung, H. Bui, and N. Ho. On cross-layer alignment for model fusion of heterogeneous neural networks. *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2023. (Top 3%)
7. K. Nguyen*, **D. Nguyen***, T. A. V. Le, T. Pham, and N. Ho. Improving mini-batch optimal transport via partial transportation. *International Conference on Machine Learning (ICML)*, 2022.
8. K. Nguyen, **D. Nguyen**, Q. Nguyen, T. Pham, H. Bui, D. Phung, T. Le, and N. Ho. On transportation of mini-batches: A hierarchical approach. *International Conference on Machine Learning (ICML)*, 2022.

Submissions

1. **D. Nguyen**, A. Payani, B. Mirzasoleiman. A Black-Box Approach to Uncertainty Quantification in LLMs: Overcoming the Limits of Semantic Entropy. 2025.

Professional services

- Reviewer at Conference on Neural Information Processing Systems (NeurIPS) 2022-2025 (Top reviewer)
- Reviewer at the International Conference on Machine Learning (ICML) 2023-2025
- Reviewer at the International Conference on Learning Representations (ICLR) 2024-2025
- Reviewer at the International Conference on Artificial Intelligence and Statistics (AISTATS) 2023-2025
- Program Committee at AAAI 2025
- Program Committee at New Frontiers in AdvML@NeurIPS2024
- Reviewer at Workshop on Spurious Correlation and Shortcut Learning @ ICLR 2025

Honors & Awards

INTERNATIONAL

- 2023 **UCLA Graduate Dean's Scholar Award**, UCLA
2017 **Toyo Top Global Scholarship A**, Toyo University
2015 **Silver medal**, 56th International Mathematical Olympiad

California, USA
Tokyo, Japan
Chiang Mai, Thailand

DOMESTIC

- 2015 **First Prize**, Vietnam Mathematical Olympiad
2014 **Second Prize**, Vietnam Mathematical Olympiad

Hanoi, Vietnam
Hanoi, Vietnam