Dang Nguyen

dangnguyen@uchicago.edu | dangng2004.github.io/ | +1 (773) 936-0857

Education

2024 - Present **Computer Science Ph.D.**, The University of Chicago.
Advisor: Prof. Chenhao Tan

2020 - 2024 **Computer Science B.S. & Mathematics B.A.**, The University of Chicago.

Selected Publications

2025 **GPT-4V Cannot Generate Radiology Reports Yet**.
Jiang, Y., Chen, C., **Nguyen, D.**, Mervak, B., & Tan, C. In Findings of the Association for Computational Linguistics (NAACL).
{evaluation of large vision-language models, prompt engineering, error analysis}

2023 **Pragmatic Radiology Report Generation**.
Nguyen, D., Chen, C., He, H., & Tan, C. In Proceedings of Machine Learning for Health (ML4H).
{vision-language modeling, finetuning LLMs, reducing hallucinations}

Professional Experience

2022 **NetApp Inc.**, Research Intern.

- Developed machine learning algorithms to detect slow disk drives.
- Optimized an algorithm based on KL divergence and improved NetApp's threshold-based detector's true positive rate from 10% to 59%.
- Used Apache Hive and SQL to create 3 novel datasets of slow disks.

Research Experience

2023 - Present **Chicago Human+AI Lab**, Student Researcher.
Race Representations are Robust to Prompt Variation in Language Model Decision-making.
Advisor: Prof. Chenhao Tan

- Designed synthetic datasets to reveal racial biases in Gemma 2B Instruct and LLaMA 3 8B Instruct's decisions in college admissions and hiring.
- Used Distributed Alignment Search to find models' race subspaces with up to 90% accuracy.
- Debiased models' decisions with representation interventions and demonstrated their efficacy over prompt engineering.

2020 - 2022 **Machine Learning for Systems Research (UCARE Lab)**, Student Researcher.
Advisor: Prof. Haryadi Gunawi

- The project aimed to develop a machine learning model for slow disk detection.
- Built a pipeline to retrieve and parse disk performance data from a vendor's database, and optimized it by 90%.
- Applied K-means and DBSCAN on 23 out of 91 pairwise correlations of data features to detect anomalous disk clusters.
- Developed an algorithm using KL divergence to detect slow disks and analyzed 50,000 disks in 3 days using 3 high-performance compute servers.

Teaching Experience

Winter 2025 **Teaching Assistant**, CMSC 25700: Natural Language Processing, The University of Chicago.

- Designed homework assignments on transformers, pre-training, and interpretability.
- Lead a tutorial session on PyTorch.