

# Hieu N. Nguyen

State College, PA

+1 814-699-3755 | hnn5071@psu.edu | nnhieu.github.io | github.com/NNHieu | linkedin.com/in/hieunn-uet | Google Scholar

## RESEARCH INTERESTS

I am interested in (i) studying the fundamental strengths and limitations of learning algorithms through large-scale empirical research, and (ii) developing next-generation intelligent systems capable of complex System 2 capabilities—such as reasoning, planning, and creativity—to advance scientific discovery and address real-world problems at scale.

## EDUCATION

### The Pennsylvania State University

PhD in Computer Science

State College, PA

Aug. 2025 - Current

- Advisor: Prof. Rui Zhang
- Research Topics: Generative AI, Reasoning, Interpretability, AI for Scientific Discovery
- Awarded *University Graduate Fellowship*

### VNU - University of Engineering and Technology

Bachelor of Computer Science - Honors Program

Hanoi, Vietnam

Aug 2022

- Thesis: An application of Neural Tangent Kernel on DEQ Models - Advisors: Dr. Hoang Thanh-Tung and Dr. Ta Viet Cuong.

## RESEARCH EXPERIENCE

### The Pennsylvania State University

Advisor: Prof. Rui Zhang

State College, PA

Aug 2025 - Current

- Evaluated the impact of quantization, distillation, and pruning on Large Reasoning Models (LRMs), demonstrating that strategically protecting just 2% of critical weights during quantization yields a 6.57% accuracy improvement over state-of-the-art (ICLR 2026).
- Analyzed the impact of training data distributions on reasoning behaviors and coverage loss across SFT and RL post-training paradigms, culminating in a first-author paper submitted to COLM 2026.

### VinUniversity - College of Engineering & Computer Science

Research Assistant

Hanoi, Vietnam

June 2023 - Aug 2025

- Developed LLM-SRBench, a rigorous evaluation framework of 239 problems designed to prevent data memorization and test true data-driven reasoning in LLMs for scientific equation discovery (Accepted as an Oral presentation at ICML 2025). - Supervisor: Prof. Khoa Doan
- Mitigated reward over-optimization in direct alignment algorithms for LLMs by designing and implementing importance sampling techniques, resulting in a publication at NeurIPS 2025. - Supervisor: Prof. Khoa Doan
- Engineered the Correlated Low-rank Structure (CoLR) framework for Federated Recommendation Systems, integrating lightweight trainable parameters and Homomorphic Encryption to reduce communication payload size by up to 93.75% with only an 8% performance trade-off (WWW 2024 Oral). - Supervisor: Prof. Kok-Seng Wong and Prof. Andrew Le

### FPT AI Center

AI Resident - Mentor: Dr. Hoang Thanh-Tung

Hanoi, Vietnam

May 2021 - June 2023

- Main research topics: Theoretical foundations of machine learning.
- Formulated a novel cosine similarity-based method for Out-of-Distribution (OOD) detection, improving model robustness and reliability for deployment (Presented at ICML 2023 Workshop).
- Formulated selective poisoning strategies for clean-label backdoor attacks under a highly constrained, zero-knowledge threat model, maximizing attack success rates using only limited, target-class data, leading to a publication at ICLR 2025.
- Diagnosed cross-class data comparisons as the root cause of Influence Function (IF) instability in deep networks; architected a novel class-based IF solution. This approach substantially boosted anomalous data detection performance in NLP tasks without introducing computational overhead (ACL 2023).

### VNU-UET

Advisors: Dr. Hoang Thanh-Tung and Dr. Ta Viet Cuong

Hanoi, Vietnam

May 2021 - June 2022

- Formulated a mathematical proof establishing a linear convergence rate for training monotone Deep Equilibrium (DEQ) models under specific initialization conditions.

## PUBLICATIONS

Why Do Reasoning Models Lose Coverage? The Role of Data and Forks in the Road

**Nguyen Ngoc-Hieu**, Parshin Shojaee, Minh Phuc Nguyen, Nan Zhang, Chandan K. Reddy, Khoa D Doan, Rui Zhang

Submitted to *Conference on Language Modeling (COLM)*, 2026

When Reasoning Meets Compression: Understanding the Effects of LLMs Compression on Large Reasoning Models

Nan Zhang, Eugene Kwek, Yusen Zhang, **Nguyen Ngoc-Hieu**, Prasenjit Mitra, Rui Zhang

ICLR, 2026, URL: <https://openreview.net/forum?id=2za3iNkwXn>

Wicked Oddities: Selectively Poisoning for Effective Clean-Label Backdoor Attacks

Nguyen Hung-Quang, **Nguyen Ngoc-Hieu**, The-Anh Ta, Thanh Nguyen-Tang, Kok-Seng Wong, Hoang Thanh-Tung, Khoa D Doan

ICLR, 2025, URL: <https://openreview.net/forum?id=1Z3C49JQVf>

Mitigating Reward Over-optimization in Direct Alignment Algorithms with Importance Sampling

Nguyen Minh Phuc, **Nguyen Ngoc-Hieu**, Duy Minh Ho Nguyen, Anji Liu, An Mai, Binh T. Nguyen, Daniel Sonntag, Khoa D Doan

NeurIPS, 2025, URL: <https://openreview.net/forum?id=1tPRj2nthL>

LLM-SRBench: A New Benchmark for Scientific Equation Discovery with Large Language Models

Parshin Shojaee\*, **Nguyen Ngoc-Hieu\***, Kazem Meidani, Amir Barati Farimani, Khoa D Doan, Chandan K. Reddy

*ICML (Oral)*, 2025, URL: <https://openreview.net/forum?id=SyQPizJVWY>

Towards Efficient Communication and Secure Federated Recommendation System via Low-rank Training

**Nguyen Ngoc-Hieu**, Tuan-Anh Nguyen, Tuan Minh Nguyen, Vu Tien Hoang, Dung D. Le, Kok-Seng Wong

*WWW (Oral)*, 2024, URL: <https://arxiv.org/pdf/2401.03748>

A Cosine Similarity-based Method for Out-of-Distribution Detection

**Nguyen Ngoc-Hieu**, Nguyen Hung-Quang, The-Anh Ta, Thanh Nguyen-Tang, Khoa D Doan, Hoang Thanh-Tung

*Spurious Correlations, Invariance, and Stability @ ICML*, 2023, URL: <https://arxiv.org/pdf/2306.14920>

Class-based Influence Functions for Error Detection

Thang Nguyen-Duc, Hoang Thanh-Tung, Quan Hung Tran, Dang Huu-Tien, **Nguyen Ngoc-Hieu**, Anh T. V. Dau, Nghi D. Q. Bui

*Association for Computational Linguistics (ACL)*, 2023, URL: <https://aclanthology.org/2023.acl-short.104/>

## SKILLS

---

**Programming** Python (NumPy, Pandas, Scikit-learn. etc.), C/C++, CUDA.

**ML frameworks** PyTorch, Jax, DeepSpeed, FSDP, verl, Unsloth.

**DevOps** Linux, Shell (Bash/Zsh), Git, Docker, AWS.