Huy Nguyen

Last update: February 02, 2026

Contact Information

`Phone:` (+1) 571 356-7068
`University Email:` huynm@utexas.edu
`Personal Email:` huynguyen.hmn@gmail.com
`Homepage:` https://huynm99.github.io/

Education

The University of Texas at Austin, Austin, TX, USA. 2022-2027
Ph.D. Candidate in Statistics at the Department of Statistics and Data Sciences.

- Advisors: Professor Nhat Ho and Professor Alessandro Rinaldo.
- GPA: 4.0/4.0

Ho Chi Minh University of Science, Ho Chi Minh City, Vietnam. 2017–2020
Bachelor of Science in Mathematics and Computer Science, Honor Program.

- Advisor: Professor Dang Duc Trong.
- GPA: 9.75/10.0 - *Summa Cum Laude*.

Research Experience

Microsoft Corporation, Redmond, WA, USA. Summer 2024
Research Intern.

- Research topics: Applications of Mixture of Experts in Large Language Models.
- Propose a method for selecting crucial attention heads in the multi-head mechanism based on the routing strategy in mixture of experts to improve the efficiency of the Large Language Models.

The University of Texas at Austin, Austin, TX, USA. 2024-Present
Graduate Research Assistant.

- Research topic: Mixture of Experts: From Theory to Applications.
- Supervisors: Professor Nhat Ho and Professor Alessandro Rinaldo.

VinAI, Hanoi, Vietnam. 2020–2022
AI Research Resident.

- Research topics: Optimal Transport theory and its applications in Domain Adaptation.
- Skill gained: Did research on Optimal Transport (Sinkhorn algorithms, Barycenter computation, etc) and applied them to study Data Shift and Label Shift problems in Domain Adaptation.

Research Interests

My research focuses on four important aspects of Mixture-of-Experts (MoE) models, including Scalability (effective sparse MoE in large language models), Heterogeneity (MoE in multi-modal learning), Efficiency (MoE in parameter-efficient fine-tuning, namely low-rank adaptation and prompt-based tuning), and Interpretability (theoretical understandings of gating mechanism and expert structures). Additionally, I am also interested in Optimal Transport problems.

Publications

26. Minh Le, Bao-Ngoc Dao, **Huy Nguyen**, Quyen Tran, Anh Nguyen, Nhat Ho. One-Prompt Strikes Back: Sparse Mixture of Experts for Prompt-based Continual Learning. *In International Conference on Learning Representations, 2026.*
25. Minh Le, Anh Nguyen, **Huy Nguyen**, Chau Nguyen, Anh Tran, Nhat Ho. Revisit Visual Prompt Tuning: The Expressiveness of Prompt Experts. *In International Conference on Learning Representations, 2026.*
24. **Huy Nguyen**, Nhat Ho**, Alessandro Rinaldo**. Convergence Rates for Softmax Gating Mixture of Experts. *IEEE Transactions on Information Theory 72(2), 1276-1304, 2026.*
23. **Huy Nguyen**, Pedram Akbarian*, Trang Pham*, Trang Nguyen*, Shujian Zhang, Nhat Ho. Statistical Advantages of Perturbing Cosine Router in Mixture of Experts. *In International Conference on Learning Representations, 2025.*

22. Nghiem Tuong Diep*, **Huy Nguyen***, Chau Nguyen, Minh Le, Duy Minh Ho Nguyen, Daniel Sonntag, Mathias Niepert, Nhat Ho. On Zero-Initialized Attention: Optimal Prompt and Gating Factor Estimation. *Proceedings of the International Conference on Machine Learning, 2025.*

21. Tuan Truong*, Chau Nguyen*, **Huy Nguyen***, Minh Le, Trung Le, Nhat Ho. RepLoRA: Reparameterizing Low-rank Adaptation via the Perspective of Mixture of Experts. *Proceedings of the International Conference on Machine Learning, 2025.*

20. Minh Le*, Chau Nguyen*, **Huy Nguyen***, Quyen Tran, Trung Le, Nhat Ho. Revisiting Prefix-tuning: Statistical Benefits of Reparameterization among Prompts. *In International Conference on Learning Representations, 2025.*

19. Fanqi Yan*, **Huy Nguyen***, Dung Le*, Pedram Akbarian, Nhat Ho**, Alessandro Rinaldo**. On Minimax Estimation of Parameters in Softmax-Contaminated Mixture of Experts. *Advances in Neural Information Processing Systems, 2025..*

18. Fanqi Yan*, **Huy Nguyen***, Dung Le*, Pedram Akbarian, Nhat Ho. Understanding Expert Structures on Minimax Parameter Estimation in Contaminated Mixture of Experts. *In International Conference on Artificial Intelligence and Statistics, 2025.*

17. **Huy Nguyen**, Nhat Ho**, Alessandro Rinaldo**. Sigmoid Gating is More Sample Efficient than Softmax Gating in Mixture of Experts. *Advances in Neural Information Processing Systems, 2024.*

16. Xing Han, **Huy Nguyen***, Carl Harris*, Nhat Ho, Suchi Saria. FuseMoE: Mixture-of-Experts Transformers for Fleximodal Fusion. *Advances in Neural Information Processing Systems, 2024.*

15. Minh Le, An Nguyen*, **Huy Nguyen***, Trang Nguyen*, Trang Pham*, Linh Van Ngo, Nhat Ho. Mixture of Experts Meets Prompt-Based Continual Learning. *Advances in Neural Information Processing Systems, 2024.*

14. **Huy Nguyen**, Nhat Ho**, Alessandro Rinaldo**. On Least Square Estimation in Softmax Gating Mixture of Experts. *Proceedings of the International Conference on Machine Learning, 2024.*

13. **Huy Nguyen**, Pedram Akbarian, Nhat Ho. Is Temperature Sample Efficient for Softmax Gaussian Mixture of Experts? *Proceedings of the International Conference on Machine Learning, 2024.*

12. **Huy Nguyen**, Pedram Akbarian, TrungTin Nguyen, Nhat Ho. A General Theory for Softmax Gating Multinomial Logistic Mixture of Experts. *Proceedings of the International Conference on Machine Learning, 2024.*

11. **Huy Nguyen**, Pedram Akbarian, Fanqi Yan, Nhat Ho. Statistical Perspective of Top-K Sparse Softmax Gating Mixture of Experts. *In International Conference on Learning Representations, 2024.*

10. **Huy Nguyen***, TrungTin Nguyen*, Khai Nguyen, Nhat Ho. Towards Convergence Rates for Parameter Estimation in Gaussian-gated Mixture of Experts. *In International Conference on Artificial Intelligence and Statistics, 2024.*

9. **Huy Nguyen**, Khai Nguyen, Nhat Ho. On Parameter Estimation in Gaussian Deviated Mixture of Experts. *In International Conference on Artificial Intelligence and Statistics, 2024.*

8. **Huy Nguyen**, TrungTin Nguyen, Nhat Ho. Demystifying Softmax Gating Function in Gaussian Mixture of Experts. *Advances in Neural Information Processing Systems, 2023 (Spotlight, Top 3.6% out of 12343 submissions).*

7. Dat Do*, **Huy Nguyen***, Khai Nguyen, Nhat Ho. Minimax Optimal Rate for Parameter Estimation in Multivariate Deviated Models. *Advances in Neural Information Processing Systems, 2023.*

6. Dung Le*, **Huy Nguyen***, Khai Nguyen*, Trang Nguyen*, Nhat Ho. Fast Approximation of the Generalized Sliced-Wasserstein Distance. *IEEE International Conference on Acoustics, Speech and Signal Processing, 2024.*

5. Khai Nguyen, Tongzheng Ren, **Huy Nguyen**, Litu Rout, Tan Nguyen, Nhat Ho. Hierarchical Sliced Wasserstein Distance. *In International Conference on Learning Representations, 2023.*

4. Khang Le*, Dung Le*, **Huy Nguyen***, Dat Do, Tung Pham, Nhat Ho. Entropic Gromov-Wasserstein between Gaussian Distributions. *Proceedings of the International Conference on Machine Learning, 2022.*

3. Khang Le*, **Huy Nguyen***, Khai Nguyen, Tung Pham, Nhat Ho. On Multimarginal Partial Optimal Transport: Equivalent Forms and Computational Complexity. *In International Conference on Artificial Intelligence and Statistics, 2022.*

2. Khang Le*, **Huy Nguyen***, Quang Minh Nguyen, Tung Pham, Hung Bui, Nhat Ho. On Robust Optimal Transport: Computational Complexity and Barycenter Computation. *Advances in Neural Information Processing Systems, 2021.*

1. Thu Nguyen, Duy H. M. Nguyen, **Huy Nguyen**, Binh T. Nguyen, Bruce A. Wade. EPEM: Efficient Parameter Estimation for Multiple Class Monotone Missing Data. *Information Sciences Journal, Volume 567, page 1-22.*

Preprints

10. Viet Nguyen*, Tuan Minh Pham*, Thinh Cao*, Tan Dinh, **Huy Nguyen**, Nhat Ho**, Alessandro Rinaldo**. A Statistical Theory of Gated Attention through the Lens of Hierarchical Mixture of Experts. *Under review, arXiv:2602.01468.*

9. Tuan Minh Pham*, Thinh Cao*, Viet Nguyen*, **Huy Nguyen**, Nhat Ho**, Alessandro Rinaldo**. Rethinking Multinomial Logistic Mixture of Experts with Sigmoid Gating Function. *Under review, arXiv:2602.01466.*

8. Fanqi Yan*, Dung Le*, Trang Pham, **Huy Nguyen**, Nhat Ho. Improving Minimax Estimation Rates for Contaminated Mixture of Multinomial Logistic Experts via Expert Heterogeneity. *Under review, arXiv:2602.00939.*

7. **Huy Nguyen**, Thong T. Doan, Quang Pham, Nghi Bui, Nhat Ho**, Alessandro Rinaldo**. On DeepSeekMoE: Statistical Benefits of Shared Experts and Normalized Sigmoid Gating. *Under review, arXiv:2505.10860.*

6. Nghiem T. Diep, Dung Le, Tuan Truong, Tan Dinh, **Huy Nguyen**, Nhat Ho. Hypernetwork-Driven Low-Rank Adaptation Across Attention Heads. *Under review, arXiv:2510.04295.*

5. Nghiem T. Diep, Hien Dang, Tuan Truong, Tan Dinh, **Huy Nguyen**, Nhat Ho. DoRAN: Stabilizing Weight-Decomposed Low-Rank Adaptation via Noise Injection and Auxiliary Networks. *Under review, arXiv:2510.04331.*

4. **Huy Nguyen***, Fanqi Yan*, Pedram Akbarian, Nhat Ho**, Alessandro Rinaldo**. Sigmoid Self-Attention has Lower Sample Complexity than Softmax Self-Attention: A Mixture-of-Experts Perspective. *Under review, arXiv:2502.00281.*

3. **Huy Nguyen***, Xing Han*, Carl Harris, Suchi Saria**, Nhat Ho**. On Expert Estimation in Hierarchical Mixture of Experts: Beyond Softmax Gating Functions. *Under review, arXiv:2410.02935.*

2. Pedram Akbarian*, **Huy Nguyen***, Xing Han*, Nhat Ho. Quadratic Gating Mixture of Experts: Statistical Insights into Self-Attention. *Under review, arXiv:2410.11222.*

1. Nam V. Nguyen, **Huy Nguyen**, Quang Pham, Van Nguyen, Savitha Ramasamy, Nhat Ho. CompeteSMoE - Statistically Guaranteed Mixture of Experts Training via Competition. *Under review, arXiv:2505.13380.*

Presentations

5. Mixture of Experts in Large-scale and Multimodal Models. *MIT reading group hosted by Prof. Priya Donti, Virtual, 2025 (Invited talk).*

4. Mixture of Experts in Large-scale and Multimodal Models. *Two Sigma PhD Fellowship Finalist Reception, Virtual, 2025 (Invited talk).*

3. Mixture of Experts in Large-scale and Multimodal Models. *Dartmouth Applied and Computational Mathematics Seminar, Virtual, 2025 (Invited talk).*

2. Demystifying Softmax Gating Function in Gaussian Mixture of Experts. *STATML@UT Reading Group, Austin, TX, 2024 (Invited talk).*

1. Demystifying Softmax Gating Function in Gaussian Mixture of Experts. *IFML Workshop on Generative AI (Student talks session), Austin, TX, 2023 (Invited talk).*

Teaching Experience

The University of Texas at Austin, Austin, TX, USA.
Teaching Assistant at the Department of Statistics and Data Sciences.

- SDS302F - Foundations of Data Analysis. Fall 2022
- SDS322E - Elements of Data Science. Spring 2023
- SDS320E - Elements of Statistics. Fall 2023
- SDS315 - Statistical Thinking. Spring 2025

Professional Services

Program Committee/Reviewer at

- the Electronic Journal of Statistics (EJS).
- the Journal of Machine Learning Research (JMLR).
- the IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI).
- the Transactions on Machine Learning Research (TMLR).
- the International Conference on Machine Learning (ICML) 2022-2026.
- the Conference on Neural Information Processing Systems (NeurIPS) 2022-2025.
- the International Conference on Artificial Intelligence and Statistics (AISTATS) 2022-2026.
- the International Conference on Learning Representations (ICLR) 2024-2026.
- the Association for the Advancement of Artificial Intelligence (AAAI) 2025-2026.

Co-organizer of the Statistical Machine Learning seminar at UT Austin (STATML@UT).

Professional Membership

- Regular Member of the Institute of Electrical and Electronics Engineers (IEEE).
- Student Member of the Institute of Mathematical Statistics (IMS).

Honors and Awards

- Top Reviewer at NeurIPS 2024. 2024
- AISTATS 2024 Registration Grant. 2024
- ICLR 2024 Travel Award. 2024
- NeurIPS 2023 Scholar Award. 2023

Technical Skills

- *System*: MacOS, Linux, Windows.
- *Programming Languages*: Python (Pytorch, Sci-kit Learn, Numpy, Matplotlib), R, MATLAB.

References

- **Nhat Ho.** Email: minhnhat@utexas.edu (Advisor).
- **Alessandro Rinaldo.** Email: alessandro.rinaldo@austin.utexas.edu (Advisor).