

# Hoang Phan

☎ 646-508-0988 | ✉ phanviethoang1512@gmail.com | 🌐 viethoang1512.github.io | 📺 hoang-pv | 🌐 VietHoang1512

## EDUCATION

---

### New York University

Ph.D., Data Science, Advisors: Qi Lei, Andrew Gordon Wilson.

New York, USA

Sept. 2023 - Present

### Hanoi University of Science and Technology

B.Sc., Computer Science, Honor program.

Hanoi, Vietnam

Aug. 2018 - Aug. 2022

## EXPERIENCE

---

### Amazon

Applied Scientist Intern

Palo Alto, USA

May. 2024 - Aug. 2024

Mentors: Colin Lockard, Yifan Gao, Rongmei Lin

- Proposed a general framework for the data mixture problem that obtains SOTA performance, speeding up the pretraining of LLMs by 35%.

### VinAI Research

Research resident

Hanoi, Vietnam

Aug. 2021 - Feb. 2023

Advisor: Trung Le

- Main research topics: *Robust machine learning, Transfer learning, Multi-task learning, Self-supervised learning.*
- Participated in a Smart City project that aims to build a kidnapping and unshielded truck detection system.

### Data Science Laboratory

Undergraduate research student

SoICT-HUST

Sep. 2019 - Aug. 2021

Advisors: Khoat Than, Huong Le Thanh, Linh Ngo Van

- Main research topics: *Probabilistic inference, Continual learning.*
- Developed a processing toolkit for the Khmer language (> 40.000 downloads and installations), supporting Asian Language machine translation. Used in Meta Research projects (NLLB, LASER, stopes), NAACL Workshop ...

### Cinnamon Inc

AI research engineer

Hanoi, Vietnam

Jan. 2020 - June. 2021

- Worked on information extraction of business documents and meta-learning, question-answering system research.
- Developed an internal usage chatbot, integrated with Slack, in support of the Human Resource department.

## PUBLICATIONS

---

- Yijun Dong\*, **Hoang Phan\***, Xiang Pan\* and Qi Lei. "Sketchy Moment Matching: Toward Fast and Provable Data Selection for Finetuning". In *Advances in Neural Information Processing Systems*, 2024. [pdf]
- Hoang Phan\***, Lam Tran\*, Quyen Tran\* and Trung Le. "Enhancing Domain Adaptation through Prompt Gradient Alignment". In *Advances in Neural Information Processing Systems*, 2024. [pdf]
- Hao Phung, Quan Dao, Trung Dao, **Hoang Phan**, Dimitris N. Metaxas, Anh Tran. "DiMSUM: Diffusion Mamba - A Scalable and Unified Spatial-Frequency Method For Image Generation." In *Advances in Neural Information Processing Systems*, 2024.
- Hoang Phan**, Andrew Gordon Wilson, Qi Lei. "Controllable Prompt Tuning for Balancing Group Distributional Robustness". *International Conference on Machine Learning*, 2024. [pdf]
- Anh Nguyen, Long Vuong, **Hoang Phan**, Toan Do, Dinh Phung and Trung Le. "Flat Seeking Bayesian Neural Networks". *Advances in Neural Information Processing Systems*, 2023. [pdf]
- Hoang Phan**, Trung Le, Trung Phung, Anh Bui, Nhat Ho and Dinh Phung. "Global-Local Regularization Via Distributional Robustness". In *International Conference on Artificial Intelligence and Statistics*, 2023. [pdf]
- Hoang Phan**, Ngoc Tran, Trung Le, Toan Tran, Nhat Ho and Dinh Phung. "Stochastic Multiple Target Sampling Gradient Descent". In *Advances in Neural Information Processing Systems*, 2022. [pdf]
- Hoang Phan\***, Anh Phan\*, Son Nguyen\*, Linh Ngo Van and Khoat Than. "Reducing Catastrophic Forgetting in Neural Networks via Gaussian Mixture Approximation". In *Advances in Knowledge Discovery and Data Mining. PAKDD. Lecture Notes in Computer Science*. Springer, 2022. [pdf]

- **Hoang Phan**, Long Nguyen, Long Nguyen, and Khanh Doan. “Matching The Statements: A Simple and Accurate Model for Key Point Analysis”. In *Proceedings of the 8th Workshop on Argument Mining* at EMNLP, 2021. [pdf]
- Tam Nguyen, Quang Pham, Linh Doan, Hoang Trinh, Anh Nguyen and **Hoang Phan**. “Contrastive Learning for Natural Language-Based Vehicle Retrieval”. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2021. [pdf]
- Quang Pham, **Hoang Phan**, Hoang Trinh, Anh Nguyen and Hoai Nguyen. “S-NLP at ICDAR 2021: Multimodal Emotion Recognition on Comics scenes”. In *International Conference on Document Analysis and Recognition Workshops*. Technical report. [pdf]

---

## PREPRINTS

- **Hoang Phan**, Lam Tran, Ngoc Tran, Nhat Ho, Dinh Phung and Trung Le. “Improving Multi-task Learning via Seeking Task-based Flat Regions”. Under review. [pdf]
- Ngoc Tran, Son Duong, **Hoang Phan**, Tung Pham, Dinh Phung and Trung Le. “Sharpness & Shift-Aware Self-Supervised Learning”. Under review. [pdf]
- Ngoc Tran, Lam Tran, **Hoang Phan**, Anh Bui, Tung Pham, Toan Tran, Dinh Phung and Trung Le. “Robust Contrastive Learning With Theory Guarantee”. Under review. [pdf]
- Quyen Tran, **Hoang Phan**, Khoat Than, Dinh Phung and Trung Le. “Continual Learning with Optimal Transport based Mixture Model”. Under review. [pdf]

(\*) denotes equal contribution

---

## NATIONAL/INTERNATIONAL PRIZE

<b>National Olympiad for Students</b>	2017 - 2018
Vietnam Mathematical Society (VMS) Olympiad for universities and high school students	
• Gold medal (2018), Silver medal (2017).	
<b>Vietnam Mathematics Olympiad (VMO)</b>	2017 - 2018
The annual national mathematics competition for high school students	
• Second prize (2018), Third prize (2017).	
• International Mathematics Olympiad team selection test contestant.	
<b>International Tournament of Towns (ITOT)</b>	2017
International mathematical Olympiad for school students	
• Bronze medal (2017).	
<b>Vietnam Institute for Advanced Study in Mathematics (VIASM) Scholarship</b>	2016 - 2018
Scholarship of the National Program for the Development of Mathematics	
<b>Vallet Scholarship</b>	2017 - 2018
Odon Vallet’s scholarships for outstanding Vietnamese students	

---

## ACHIEVEMENTS

<b>ArgMining: Key Point Analysis Shared Task, EMNLP</b>	2021
Presented the 4 <sup>th</sup> solution in the workshop, hosted by IBM Research AI team.	
<b>NVIDIA AI City Challenge: Natural Language-Based Vehicle Retrieval, CVPR</b>	2021
Got the 4 <sup>th</sup> place with a Siamese model capable of extracting visual concepts from natural language.	
<b>Multimodal Emotion Recognition on Comics scenes, ICDAR</b>	2021
Won the first prize with a late fusion model, combined the processing of multiple input modalities.	
<b>Reliable Intelligence Identification on Vietnamese SNSs, VLSP</b>	2020
Achieved a competitive performance (ranked 4 <sup>th</sup> ) in the VLSP shared task.	
<b>Credit Scoring Challenge, Kalapa. Jsc</b>	2020
Team ranked 1/146 with 800 contestants in constructing a banking credit scoring system.	
<b>University of Liverpool - Ion Switching, Kaggle</b>	2020
Bronze medalist (solo) in the University of Liverpool’s Institute of Ageing and Chronic Disease competition.	
<b>Practical deep learning, Vietnam Institute for Advanced Study in Mathematics</b>	2019
Developed a deep learning-based Vietnamese spell corrector and ranked 1/50 in the final-project defense.	

---

## PROFESSIONAL SERVICES

Reviewer at NeurIPS 2023, ICLR 2024, ICCV 2023, ECCV 2024, WACV 2025.