

# Xuan Phi Nguyen

🌐 [linkedin.com/in/xuanphinguyen](https://www.linkedin.com/in/xuanphinguyen) 🌐 [nxphi47.github.io](https://github.com/nxphi47) ✉️ [nxphi47@gmail.com](mailto:nxphi47@gmail.com)

## EXPERIENCE

---

### Salesforce AI Research

Research Scientist

United States

May 2024 – Present

- Lead efforts in SFR-RAG - smaller but powerful language models specialized in Retrieval Augmented Generation (RAG). SFR-RAG is praised by CEO Marc Benioff and receives great media coverage. Conduct research in improving reliability and reasoning abilities of large language models.

### DAMO Academy - Alibaba Group

Senior Algorithm Engineer (AI Research)

Singapore

April 2023 – May 2024

- Led research and technical efforts in SeaLLMs - The first and state-of-the-art large language models for Southeast Asian languages, earning significant attentions from media and research communities.

### Facebook AI Research (FAIR), Meta

AI Research Intern

United States

May 2022 – Sep 2022

- Researched on improving Speech-to-Speech translation by introducing novel strategies to generate and augment synthetic data. Our paper is published ICASSP 2023.

### Facebook AI Research (FAIR), Meta

AI Research Intern

United States (Remote)

May 2021 – Sep 2021

- Developed a novel pseudo-parallel data mining technique and integration strategy to achieve the state of the art in Unsupervised Machine Translation - translation without human annotations. Our paper is published in ICLR 2022.

### Salesforce AI Research Asia

AI Research Intern

Singapore

May 2019 – Aug 2019

- Researched on novel neural architectures to improve various natural language processing tasks. Proposed new state-of-the-art methods and submitted papers to machine learning conferences. Our paper is published in ICLR 2020.

### Visa Inc.

Software Engineer Intern

Singapore

May 2017 – Jul 2017 and Jan 2018 – Jul 2018

- Developed a lightweight character-level convolutional neural network to perform scripted text classification tasks at up to 98% accuracy while consuming 1000 times less resources.

### Panasonic Research and Development Center

Software Engineer Intern

Singapore

May 2016 – Jul 2016

- Cooperated to develop a new AI algorithm based on Support Vector Machine to classify electrical signals, achieving 94% of experimental accuracy and 83.8% real-time accuracy.

## EDUCATION

---

### Nanyang Technological University

Doctor of Philosophy in Computer Science & Artificial Intelligence.

Singapore

Graduation: April 2023

- Thesis title: “Improving Neural Machine Translation: Data Centric Approaches”.

### Nanyang Technological University

Bachelor in Electrical & Electronics Engineering. **First Class**

Singapore

Graduation: July 2019

- Dean's List Awards – Top 5% of cohort on academic performance, AY 2016/2017 - AY 2017/2018.

## RESEARCH PUBLICATIONS

---

- [Xuan-Phi Nguyen](#), Shrey Pandit, Senthil Purushwalkam, Austin Xu, Hailin Chen, Yifei Ming, Zixuan Ke, Silvio Savarese, Caiming Xong & Shafiq Joty (2024). *SFR-RAG: Towards Contextually Faithful LLMs*. Technical Report.
- [Xuan-Phi Nguyen](#), Wenxuan Zhang, Xin Li, Mahani Aljunied, Qingyu Tan, Liying Cheng, Guanzheng Chen, Yue Deng, Sen Yang, Chaoqun Liu, Hang Zhang & Lidong Bing (2024). *SeaLLMs - Large Language Models for Southeast Asia*. In Proceedings of the Association for Computational Linguistics (**ACL 2024**), Demo Track.
- [Xuan-Phi Nguyen](#), Sharifah Mahani Aljunied, Shafiq Joty & Lidong Bing (2024). *Democratizing LLMs for Low-Resource Languages by Leveraging their English Dominant Abilities with Linguistically-Diverse Prompts*. In Proceedings of the Association for Computational Linguistics (**ACL 2024**).
- Chenhui Shen, Liying Cheng, [Xuan-Phi Nguyen](#), Yang You & Lidong Bing (2023). *Large Language Models are Not Yet Human-level Evaluators for Abstractive Summarization*. The 2023 Conference on Empirical Methods in Natural Language Processing (**EMNLP 2023**) - Findings.

- Chenhui Shen, Liying Cheng, Xuan-Phi Nguyen, Yang You & Lidong Bing (2023). *A Hierarchical Encoding-decoding Scheme for Abstractive Multi-document Summarization*. The 2023 Conference on Empirical Methods in Natural Language Processing (**EMNLP 2023**) - Findings.
- Xuan-Phi Nguyen, Sravya Popuri, Changhan Wang, Yun Tang, Ilya Kulikov & Hongyu Gong (2023). *Improving Speech-to-Speech Translation Through Unlabeled Text*. 2023 IEEE International Conference on Acoustics, Speech, and Signal Processing (**ICASSP 2023**).
- Xuan-Phi Nguyen, Shafiq Joty, Wu Kui & Aw Ai Ti (2022). *Refining Low-Resource Unsupervised Translation by Language Disentanglement of Multilingual Model*. 36th Conference on Neural Information Processing Systems (**NeurIPS 2022**).
- Xuan-Phi Nguyen, Hongyu Gong, Yun Tang, Changhan Wang, Philipp Koehn & Shafiq Joty (2022). *Contrastive Clustering to Mine Pseudo Parallel Data for Unsupervised Translation*. In International Conference on Learning Representations (**ICLR 2022**).
- Xuan-Phi Nguyen, Shafiq Joty, Thanh-Tung Nguyen, Wu Kui, & Ai Ti Aw (2021). *Cross-model Back-translated Distillation for Unsupervised Machine Translation*. In Proceedings of the 38th International Conference on Machine Learning (**ICML 2021**).
- Xuan-Phi Nguyen, Shafiq Joty, Wu Kui & Aw Ai Ti (2020). *Data Diversification: A Simple Strategy For Neural Machine Translation*. 34th Conference on Neural Information Processing Systems (**NeurIPS 2020**).
- Xuan-Phi Nguyen, Shafiq Joty, Steven Hoi & Richard Socher (2020). *Tree-Structured Attention with Hierarchical Accumulation*. In International Conference on Learning Representations (**ICLR 2020**).
- Thanh-Tung Nguyen, Xuan-Phi Nguyen, Shafiq Joty & Xiaoli Li (2021). *A Conditional Splitting Framework for Efficient Constituency Parsing*. In Proceedings of Association for Computational Linguistics (**ACL 2021**).
- Thanh-Tung Nguyen, Xuan-Phi Nguyen, Shafiq Joty & Xiaoli Li (2021). *RST Parsing from Scratch*. In the North American Chapter of the Association for Computational Linguistics (**NAACL 2021**).
- Xuan-Phi Nguyen\*, Thanh-Tung Nguyen\*, Shafiq Joty & Xiaoli Li (2020). *Differentiable Window for Dynamic Local Attention*. In Proceedings of the Association for Computational Linguistics (**ACL 2020**). Co-first-authored.
- Thanh-Tung Nguyen, Xuan-Phi Nguyen, Shafiq Joty & Xiaoli Li (2020). *Efficient Constituency Parsing by Pointing*. In Proceedings of the Association for Computational Linguistics (**ACL 2020**).
- Xuan-Phi Nguyen & Shafiq Joty (2019). *Phrased-based Attentions*. arXiv preprint arXiv:1810.03444.
- P. X. Nguyen, Z. Lu, W. Huang, S. Huang, A. Katsuki & Z. Lin (2019). *Medical image segmentation with stochastic aggregated loss in a unified U-Net*. In 2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI) (**IEEE BHI 2019**), Chicago, USA.

## HONORS & AWARDS

---

### Reviewers of Artificial Intelligence Publication Conferences

- NeurIPS 2021, 2022, 2023, 2024; ICLR 2022,2023,2024; ICML 2023,2024

### 2021 Singapore Data Science Consortium (SDSC) Dissertation Research Award

Jan - Dec 2022

- Awarded for innovative and impactful thesis work of PhD students whose research interests are in data science.

### NeurIPS Student Scholar Award

Nov 2022

- Awarded for PhD students with first-authored papers at NeurIPS 2022 conference at New Orleans, USA.

### A\*STAR Computing and Information Science (ACIS) Scholarship

Jul 2019 – Jul 2023

- Fully funded by A\*STAR to pursue the doctorate program in NTU, Singapore.

### The Global Undergraduate Awards - Highly Commended Paper

Sep 2019

- Top 10% out of more than 3500 papers submitted globally.

### Shopee - IET Machine Learning Challenge Competition - 2<sup>nd</sup> Runner-up

Mar 2018

- Designed ensemble techniques to achieve 2<sup>nd</sup> Runner-up prize with 85.662% accuracy in image classification task.

### SUTD What The Hack Hackathon - Best Software Hack - Best Security Hack

Sep 2017

- Eye Gesture Control Application: VR device powered by neural networks to help the disabled to perform daily tasks.

### Singapore Scholarship

Jul 2015 – Jul 2019

- Fully funded by Ministry of Foreign Affairs of Singapore to pursue the undergraduate program in NTU, Singapore.