# THINH PHAM

⚇ Google Scholar | ✉ thinhphp@vt.edu | 🌐 thinhphp.github.io

Blacksburg, VA, USA, 24060

## RESEARCH INTERESTS

I am particularly interested in designing effective and efficient methods to keep LLMs up-to-date with dynamic real-world knowledge. Currently, my research focuses on **search-augmented LLMs** and **web search agents**.

## EXPERIENCE

- **Intelligent Internet [🌐]** *Jun 2025 - Aug 2025*
  *Research Intern*
  ◦ Supervisor: Duy Phung
  ◦ Research topic: Self-evolving agents
  ◦ Learned and researched methods to improve agents that employ dynamic self-modification to continuously refine their behaviors and decision-making strategies

- **VinAI Research (acquired by Qualcomm) [🌐]** *Jul 2022 - Jul 2024*
  *Research Resident*
  ◦ Supervisor: Dr. Dat Quoc Nguyen - Head of NLP Department.
  ◦ Research topic: Spoken language understanding
  ◦ Conducted research across a wide range of NLP tasks, leading a project from idea proposal to publication at top-tier conferences

- **VinBrain (acquired by NVIDIA) [🌐]** *Jun 2021 - Sep 2021*
  *Applied Scientist Intern*
  ◦ Topic: Automated Speech Recognition
  ◦ Optimized ASR system by implementing beam search decoding to reduce inference time and WER score

## EDUCATION

- **Virginia Tech** *Aug 2024 - now*
  *PhD in Computer Science and Applications* Virginia, USA
  ◦ Research focus: search-augmented LLMs, web search agents
  ◦ Advisor: Prof. Tu Vu

- **University of Science, Vietnam National University** *Aug 2018 - Nov 2024*
  *Bachelor's degree in Computer Science (Advanced Program)* Ho Chi Minh city, Vietnam
  ◦ GPA: 3.92/4.0
  ◦ Thesis: Integrating Label Attention into CRF-based Vietnamese Constituency Parser
  ◦ Advisor: Prof. Dien Dinh

## PREPRINTS AND PUBLICATIONS  C=CONFERENCE, P=PREPRINT

**[P]** **Thinh Pham**, Nguyen Nguyen, Pratibha Zunjare, Weiyuan Chen, Yu-Min Tseng, Tu Vu (2025). **SealQA: Raising the Bar for Reasoning in Search-Augmented Language Models**. *arXiv:2506.01062*
// Our benchmark dataset has been used by Google's Gemini, DeepSeek, and Kimi.

**[C]** Linh The Nguyen, **Thinh Pham**, Dat Quoc Nguyen (2023). **XPhoneBERT: A Pre-trained Multilingual Model for Phoneme Representations for Text-to-Speech**. In *Proceedings of the Annual Conference of the International Speech Communication Association (INTERSPEECH)*.

**[C]** **Thinh Pham**, Dat Quoc Nguyen (2023). **JPIS: A Joint Model for Profile-based Intent Detection and Slot Filling with Slot-to-Intent Attention**. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*.

**[C]** **Thinh Pham**, Chi Tran, Dat Quoc Nguyen (2023). **MISCA: A Joint Model for Multiple Intent Detection and Slot Filling with Intent-Slot Co-Attention**. In *Findings of Empirical Methods in Natural Language Processing (EMNLP)*.

## SKILLS

- **Deep Learning Framework:** PyTorch, TensorFlow, AutoGen, DSPy
- **LLMs & NLP:** Prompt engineering, fine-tuning, RAG, self-improvement and tool-augmented agents
- **Evaluation:** Benchmarking LLMs, efficiency/performance profiling, human evaluation, experimental design
- **Research:** Experimental design, literature review, interdisciplinary collaboration, academic writing and presentation

## Honors and Awards

- Virginia Tech Graduate Assistantships *2024 - 2025*
- A number of academic scholarships for undergraduate students *2018 - 2022*
- Second Prize, National Olympiad in Informatics, Vietnam *2017*
- Second Prize, National Informatics Contest for Youth, Vietnam *2017*

## Projects

- **Named Entity Recognition** *2021*
  *The International Workshop on Vietnamese Language and Speech Processing* [🌐]
  - Built machine learning models to recognize entities in Vietnamese documents
  - Handled nested entities with a span-based model, and noisy labels from the data
  - Attained the $2^{nd}$ rank and presented at the workshop

- **Emotions and Themes in Music** *2021*
  *The Multimedia Evaluation Workshop* [🌐]
  - Automatically recognized the emotions and themes conveyed in a music recording using machine learning algorithms
  - Ensembled different backbone models with a co-teaching training strategy
  - Attained the $4^{th}$ rank and presented at the workshop

## Professional Responsibilities

- **Graduate Teaching Assistant** *2024 - 2025*
  *Virginia Tech*
  - CS3114 - Data Structure and Algorithms (*Spring 2025*)
  - CS1064 - Intro to Python Programming (*Fall 2025*)

- **Peer Review**: NAACL (2024)