Viet Hoang Pham

viethoangpham.contact@gmail.com (+84) 377 257 482 Hoang Pham

OVERVIEW

Graduated from Hanoi University of Science and Technology with a focus on Natural Language Processing. Over the past two years, at Viettel AI, I have conducted research on interpretable and trustworthy NLP systems that address large language models' challenges, including outdated knowledge, inconsistent reasoning, limited explainability and multi-modality. My recent work focuses on enhancing knowledge-intensive tasks by integrating large language models with external knowledge sources, such as knowledge graphs and textual corpora, and extending this approach to additional modalities including images, audio, and video.

EDUCATION

Hanoi University of Science and Technology *Hanoi, Vietnam*
B.Sc. in Computer Science *Sep. 2020 – Sep. 2024*

- Overall GPA: 3.75/4.0 (9.38/10.0 equivalent), top 5% of major
- Excellent Student title in all Academic Years.
- Thesis: Continual Learning and Task Retrieval for Fine-tuning Large Language Models - Score: 9.3/10.0.
- Language: English - Professional working proficiency (IELTS 7.5 Overall)

RESEARCH EXPERIENCE

- **Hoang Pham***, Thanh-Do Nguyen*, & Khac-Hoai Nam Bui (2025). "ClaimPKG: Enhancing Claim Verification via Pseudo-Subgraph Generation with Lightweight Specialized LLM." Findings of the *Annual Meeting of the Association for Computational Linguistics* ***(ACL 2025)*** **(CORE A*)**.
  Link: https://aclanthology.org/2025.findings-acl.274
- **Hoang Pham**, Thanh-Do Nguyen, & Khac-Hoai Nam Bui (2025). "Verify-in-the-Graph: Entity Disambiguation Enhancement for Complex Claim Verification with Interactive Graph Representation." In *Proceedings of the Annual Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics* ***(NAACL 2025)*** **(CORE A)** [In-person presentation].
  Link: https://aclanthology.org/2025.naacl-long.268
- Quang-Vinh Nguyen, Quang-Chieu Nguyen, **Hoang Pham**, & Khac-Hoai Nam Bui (2025). "Spec-TOD: A Specialized Instruction-Tuned LLM Framework for Efficient Task-Oriented Dialogue Systems." Accepted at the *Meeting of the Special Interest Group on Discourse and Dialogue* ***(SIGDIAL 2025)***.
  Link: https://arxiv.org/abs/2507.04841
- **Hoang Pham**, Thuy-Duong Nguyen, & Khac-Hoai Nam Bui (2024). "Agent-UniRAG: A Trainable Open-Source LLM Agent Framework for Unified Retrieval-Augmented Generation Systems" (Preprint, 2024).
  Link: https://arxiv.org/abs/2505.22571

WORKING EXPERIENCE

Artificial Intelligence Researcher *Hanoi, Vietnam*
Viettel AI, Viettel Group *Nov. 2023 – Present*

- **Conducted applied research** on high-impact topics at the intersection of natural language processing, reasoning and knowledge augmentation, with an emphasis on:
  1. **Automated Fact Checking:** Proposed reliable, interpretable frameworks integrating large language models (LLMs) with knowledge sources (e.g., knowledge graphs, raw-text corpora) for fact verification.
  2. **Agentic Reasoning Systems:** Designed scalable, interpretable agent-based frameworks for complex NLP tasks, including multi-hop question answering and claim verification.
  3. **Knowledge Graphs:** Proposed novel methodologies for constructing and exploiting knowledge graphs, with a focus on ensuring completeness, reliability, and interpretability.
  4. **Reasoning Models:** Trained specialized models based on reinforcement learning algorithms (e.g., GRPO and PPO) to enhance reasoning capability and task-specific performance.

- **Continual pre-trained large language models (LLMs) for Vietnamese:** Spearheaded infrastructure configuration, hyperparameter optimization, training script development, and large-scale data preparation (several GBs of text) to produce LLMs ranging from 7B to 70B parameters, achieving high performance in both general Vietnamese language understanding and domain-specific legal applications.
- **Aligned LLMs to desired behaviors via synthetic data:** Curated and validated Vietnamese instruction-following datasets using large-scale models (up to 405B parameters). Leveraged knowledge distillation and agentic pipelines to generate high-quality, behavior-aligned data for fine-tuning.
- **Developed AI-powered NLP systems for internal deployment:** Designed end-to-end pipelines for NER and virtual assistants, including data augmentation, model training, and deployment. Developed instruction-tuned models and RAG workflows for accurate, efficient document-based responses.

Viettel Digital Talent Residency *Hanoi, Vietnam*
Viettel Group *Apr. 2023 – Nov. 2023*

- **Intensive trained on key areas of Data Science & AI**, covering advancements in computer vision, natural language processing, speech processing, and big data management.
- **Independent mini-research improved virtual assistant performance in public services**, leveraging LLM agent techniques. Hands-on experiences in training and deploying medium-scale language models.

HONORS & AWARDS

- Top-4 Global Public, AI City Challenge 2024 – Safety Description and Analysis Track.
  Utilized grounding information with Vision-Language models to enhance factual accuracy in video description.
- Talent Ambassador (Finalist Top 10), Viettel Digital Talent Residency 2023.
- Third Prize, Hanoi Municipal Mathematics Olympiad – Senior Year of High School.

KNOWLEDGE & SKILLS

Technical Domains: Natural Language Processing (NLP), Machine Learning, Deep Learning, Computer Vision, Knowledge Graphs, Agent-based Reasoning, Reinforcement Learning.

Research Skills: Scientific Writing, Literature Review, Experimental Design, Ablation Studies, Dataset Curation & Validation.

Theoretical Foundations: Probability & Statistics, Linear Algebra, Optimization, Calculus, Information Theory, Data Structures & Algorithms.

Frameworks: PyTorch, Tensorflow, Hugging Face.
Programming Languages: Python, C++.