

LE THIEN PHUC NGUYEN

(608)-213-9350 | plnguyen2908@gmail.com | [plnguyen2908.github.io](https://github.com/plnguyen2908) | [linkedin.com/in/plnguyen2908/](https://www.linkedin.com/in/plnguyen2908/) | Google Scholar

EDUCATION

University of North Carolina at Chapel Hill

Aug 2026 - May 2031

Ph.D. in Computer Science

Advisor: **Zhongzheng (Jason) Ren**

Co-Advisor: **Yong Jae Lee**

University of North Carolina at Chapel Hill

Aug 2026 - May 2028

M.S. in Computer Science

University of Wisconsin-Madison

Sep 2022 - May 2026

B.S. in Computer Science, Data Science, Mathematics, and Statistics

GPA: 3.97/4.0

Graduated with Distinction & Dean's Honor List

Advisor: **Yong Jae Lee**

CURRENT RESEARCH PROJECTS

Improving fine-grained visual reasoning in SAM-Audio

2026 - Present

- Developing a reasoning dataset to develop the fine-grained reasoning capability for SAM-Audio.

Long-Horizon Cross-Episode Memory for Mobile Manipulation

2026 - Present

- Developing memory mechanisms for mobile manipulation agents to retain and retrieve information across episodes, enabling reasoning and action over long-horizon tasks.

PUBLICATIONS AND MANUSCRIPTS

* denotes equal contribution

- **See, Hear, and Understand: Benchmarking Audiovisual Human Speech Understanding in Multimodal Large Language Models**
Le Thien Phuc Nguyen*, Zhuoran Yu*, Samuel Low Yu Hang, Subin An, Jeongkik Lee, Yohan Ban, SeungEun Chung, Thanh-Huy Nguyen, Juwan Maeng, Soochahn Lee, and Yong Jae Lee
Findings of The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR Findings), 2026
- **LASER: Lip Landmark Assisted Speaker Detection for Robustness**
Le Thien Phuc Nguyen*, Zhuoran Yu*, and Yong Jae Lee
The IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2026
Oral.
- **Revisiting Active Speaker Detection: An In-the-Wild Benchmark for Generalization and Robustness**
Le Thien Phuc Nguyen*, Zhuoran Yu*, Khoa Quang Nhat Cao, Yuwei Guo, Tu Ho Manh Pham, Tuan Tai Nguyen, Toan Ngo Duc Vo, Lucas Poon, Soochahn Lee, and Yong Jae Lee
Interspeech, 2026
Oral.
- **DocHop: Benchmarking Out-of-domain Multi-hop Reasoning in Information-Dense Documents**
Zhuoran Yu, Le Thien Phuc Nguyen, Jaden Park, Xinyi Gu, Zexue He, Soochahn Lee, Rogerio Feris, and Yong Jae Lee
International Conference on Machine Learning (ICML), 2026
- **HARPS: Mitigating Multimodal Hallucinations via Head-Selective and Adaptive Perceptual Sink Reweighting**
Le Thien Phuc Nguyen, Zhuoran Yu, Zhongzheng (Jason) Ren, and Yong Jae Lee
Preprint, 2026
- **QwenVLConnector: A Fast, Unified Medical VLM Chatbot for Fine-Grained Clinical Perception and Text Generation**
Le Thien Phuc Nguyen*, Thien Nguyen*, Thanh-Huy Nguyen, Gia Minh Hoang, Anh Mai Vu, Ulas Bagci
FLARE Workshop @ International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2025
Oral.

- **Describe Anything Model for Visual Question Answering on Text-rich Images**
Yen-Linh Vu, Dinh-Thang Duong, Truong-Binh Duong, Anh-Khoi Nguyen, Thanh-Huy Nguyen, **Le Thien Phuc Nguyen**, Jianhua Xing, Xingjian Li, Tianyang Wang, Ulas Bagci, Min Xu
VisionDocs Workshop @ the International Conference on Computer Vision (ICCV), 2025
- **GMAT: Grounded Multi-Agent Clinical Description Generation for Text Encoder in Vision-Language MIL for Whole Slide Image Classification**
Ngoc Bui Lam Quang, Nam Le Nguyen Binh, Thanh-Huy Nguyen, **Le Thien Phuc Nguyen**, Quan Nguyen, and Ulas Bagci
ELAMI Workshop @ International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2025

EXPERIENCE

- RAIR Lab, Department of Computer Science - UNC Chapel Hill** Aug 2026 - Present
Research Assistant. *Advised by:* Prof. Zhongzheng (Jason) Ren Chapel Hill, NC
- Conducting research in embodied AI and robot learning, with a focus on long-horizon memory and mobile manipulation.
 - Developing methods that enable embodied agents to retain, retrieve, and reuse information across episodes for long-horizon tasks.
- Wisconsin AI Vision Lab, Department of Computer Science - UW-Madison** June 2024 - Aug 2026
Research Assistant. *Advised by:* Prof. Yong Jae Lee Madison, WI
- Conducted extensive literature review in the following fields: Active Speaker Detection, Lip Reading Audio-to-audio Translation Vision Language Models, and VLM Benchmarks.
 - Came up a novel method utilizing lip landmark to distill knowledge to Active Speaker Detection models through a consistency loss which helped achieve robust performance on high-noise environment.
 - Led a team of 10 UW-Madison students to conduct the annotation for the first challenging dataset in the Active Speaker Detection field which improves models performance on real word scenarios.
 - Conducted analysis on the performance of both proprietary and open multimodal large language models on understanding human conversation audiovisually.
 - Participated in writing papers and doing rebuttal during paper submission.

AWARDS

- | | |
|--|------|
| ICPC North Central North America (NCNA) Regional Contest, Gold Medal (Rank 4/116) | 2024 |
| ICPC North Central North America (NCNA) Regional Contest, Silver Medal (Rank 5/116) | 2023 |
| ICPC Vietnam National Round Contest, Second Prize | 2022 |
| Vietnam National Olympiad in Informatics, Third Prize | 2022 |

TECHNICAL SKILLS

Programming: Python, C++, Go, R
ML Frameworks: PyTorch, Hugging Face Transformers, NumPy, Scikit-learn
Research: Multimodal LLMs, Vision-Language Models, Audio-Visual Learning, Hallucination & Interpretability, Embodied AI, Mobile Manipulation, Long-Horizon Memory
Systems & Tools: Linux, Git, Docker, HTCondor, Weights & Biases