Ngoc Son Nguyen

nngocson01042002@gmail.com

nngocson — Google Scholar — nngocson2002

RESEARCH INTERESTS

Multimodal AI, Generative AI, Vision-Language Models, Speech Processing, Audio-Visual Speech Generation

EDUCATION

University of Science - Vietnam National University HCMC (HCMUS-VNU), Vietnam — Sep 2020 - Sep 2024
Bachelor of Science in Data Science

Graduated with **Highest Distinction** (GPA: 9.27/10.00, 3.86/4.00 – Ranked **2nd** in the Class of 2024)
Thesis: Advancing Vietnamese Visual Question Answering with Transformer and Convolutional Integration (Score: **10.0/10.0**)
Thesis Supervisor: Dr. Tung Le

- University Scholarship for Outstanding Academic Performance (Semesters 1–5).
- Certificate of Achievement for Outstanding Research Achievement, University of Science – Vietnam National University, Ho Chi Minh City (HCMUS-VNU).

WORK EXPERIENCE

AI Center, FPT Software — Ho Chi Minh, Vietnam
AI Research Resident — Aug 2024 - Present

Supervisor: Dr. Van Nguyen
Research topics: Multimodal AI, Speech processing, Medical imaging

- **Project:** *Design a speech synthesis model optimized for smaller model size, and faster inference speed*
  Contributed to the idea of using a learned prior as initialization instead of starting from Gaussian noise, followed by a rectified flow transformation to reach the target distribution, demonstrating improved effectiveness. Implemented the Montreal Forced Aligner to align phonemes with audio, and conducted experiments that yielded valuable insights into emerging challenges.
- **Project:** *Design a novel vision-language model for Chest X-ray localization and classification*
  Proposed a novel vision-language model tailored for medical imaging, specifically for joint localization and classification tasks on chest X-rays. The unified framework integrates multiple pre-trained visual encoders, such as BiomedCLIP and PubMedCLIP, with LLaMA 2 to enhance diagnostic accuracy. Established a strong baseline for future research, conducted ablation studies, and performed comprehensive result analysis.

Multi-modal Machine Comprehension Group, FIT-HCMUS — Ho Chi Minh, Vietnam
Research Assistant — Sep 2023 - Jun 2024

Research topics: Modeling multimodal knowledge.

- Joined a research group focused on multi-modal information mining and processing problems.
- Developing deep learning models and conducting experiments to address the Visual Question Answering (VQA) problem.
- Integrate novel methodologies, enhancing the accuracy and efficiency of VQA systems.
- Proposed the integration of Transformer architectures with Convolutional Neural Networks to improve image information extraction. Leveraged pre-trained Vietnamese language models, such as PhoBERT and BARTpho, to enhance linguistic understanding. Studied effective fusion techniques for visual and textual information, and conducted extensive research to optimize model performance on the ViVQA dataset, achieving state-of-the-art results. Designed and executed ablation studies, accompanied by thorough analysis of experimental outcomes.

PUBLICATIONS & PREPRINTS

(*): Equal contribution

[1] **Ngoc-Son Nguyen**, Thanh V. T. Tran, Hieu-Nghia Huynh-Nguyen, Truong-Son Hy, Van Nguyen. DiFlowDubber: Discrete Flow Matching for Automated Video Dubbing via Cross-Modal Alignment and Synchronization, (*Under review*)
[2] **Ngoc-Son Nguyen**, Thanh V. T. Tran, Jeongsoo Choi, Hieu-Nghia Huynh-Nguyen, Truong-Son Hy, Van Nguyen. DiFlow-TTS: Compact and Low-Latency Zero-Shot Text-to-Speech with Factorized Discrete Flow Matching, (*Under review*)
[3] Thanh V. T. Tran, **Ngoc-Son Nguyen**, Luong Tran, Long-Khanh Pham, Paarth Neekhara, Shehzeen Samarah Hussain, Van Nguyen. Precise Video-to-Audio Generation with Cross-Modal Alignment in Latent Space, (*Under review*)
[4] Hieu-Nghia Huynh-Nguyen, Huynh Nguyen Dang, **Ngoc-Son Nguyen**, Van Nguyen. Flamed-TTS: Flow Matching Attention-Free Models for Efficient Generating and Dynamic Pacing Zero-shot Text-to-Speech, (*Under review*)
[5] Cuong Tran Van*, Trong-Thang Pham*, **Ngoc-Son Nguyen**, Duy Minh Ho Nguyen, Ngan Le. DuFal: Dual-Frequency-Aware Learning for High-Fidelity Extremely Sparse-view CBCT Reconstruction, *Transactions on Machine Learning Research* (TMLR 2025)
[6] Trong-Thang Pham, AKASH AWASTHI, Saba Khan, Esteban Duran Marti, Tien-Phat Nguyen, Khoa Vo, Minh Tran, **Ngoc-Son Nguyen**, Cuong Tran Van, Yuki Ikebe, Anh Totti Nguyen, Anh Nguyen, Zhigang Deng, Carol C. Wu, Hien Van Nguyen, Ngan

Le. CT-ScanGaze: A Dataset and Baselines for 3D Volumetric Scanpath Modeling , *Proceedings of the IEEE/CVF International Conference on Computer Vision 2025 (ICCV 2025) (**Highlight**)*

[7] Hieu-Nghia Huynh-Nguyen, **Ngoc-Son Nguyen**, Huynh Nguyen Dang, Thieu Vo, Truong-Son Hy, Van Nguyen. OZSpeech: One-step Zero-shot Speech Synthesis with Learned-Prior-Conditioned Flow Matching , *Proceedings of the 63nd Annual Meeting of the Association for Computational Linguistics (ACL 2025)*

[8] Khai Le-Duc*, Ryan Zhang*, **Ngoc-Son Nguyen***, Tan-Hanh Pham, Anh Dao, Ba Hung Ngo, Anh Totti Nguyen, Truong-Son Hy. LiteGPT: Large Vision-Language Model for Joint Chest X-ray Localization and Classification Task, (*Preprint*)

[9] **Ngoc-Son Nguyen**, Van Son Nguyen, Tung Le. Advancing Vietnamese Visual Question Answering with Transformer and Convolutional Integration, *Computers and Electrical Engineering (Q1, IF = 4.9)*

TECHNICAL SKILLS

- **Programming:** C/C++, Python.
- **Framework:** PyTorch, Pytorch Lightning, SQL, Scikit-learn, Pandas, Matplotlib.
- **Tools**: Linux, Git/GitHub, LaTeX.
- **Languages**: Vietnamese (Native), English (740/990 TOEIC Listening and Reading; 310/400 TOEIC Speaking and Writing).

REFERENCES

Dr. Van Nguyen
AI LAB Manager
FPT Software AI Center, Vietnam
E-mail: vannth19@fpt.com
Scholar Profiles: Google Scholar

Prof. Truong-Son Hy
Tenure-Track Assistant Professor
Department of Computer Science & Heersink School of Medicine
The University of Alabama at Birmingham, United States
E-mail: thy@uab.edu
Scholar Profiles: Personal Page —Google Scholar

Dr. Tung Le
Lecturer
Department of Knowledge Engineering, Faculty of Information Technology
University of Science - Vietnam National University HCMC, Vietnam
E-mail: lttung@fit.hcmus.edu.vn
Scholar Profiles: Google Scholar

Dr. Huy Tien Nguyen
Lecturer
Department of Computer Science, Faculty of Information Technology
University of Science - Vietnam National University HCMC, Vietnam
E-mail: ntienhuy@fit.hcmus.edu.vn
Scholar Profiles: Google Scholar