Nhut Minh Nguyen

Ho Chi Minh City, Vietnam | minhnhut.ngnn@gmail.com | +84 939 323 874

Personal Website | GitHub | LinkedIn | Google Scholar

Summary and Objective

Nguyen Minh Nhut is an Artificial Intelligence undergraduate and Research Assistant at FPT University, specializing in multimodal emotion recognition and human-centered AI. He builds robust audio–text–vision systems using Transformer encoders, graph-based learning, and semi-supervised/federated paradigms to improve generalization under limited labels and privacy constraints, delivering reproducible pipelines and deployable inference workflows.

Education

FPT University (Ho Chi Minh City), B.Sc. in Artificial Intelligence — Sep 2022 – May 2026 (Expected)

- **Focus:** Machine Learning, Deep Learning, Speech Processing, Computer Vision, Human–Computer Interaction.
- **Relevant Coursework:** Machine Learning, Computer Vision, Speech Processing, Data Structures & Algorithms.
- **GPA:** 8.45.

Experience

Research Assistant, AiTA Lab – FPT University — Mar 2023 – Present

- Research and develop multimodal emotion recognition systems with Transformer encoders, feature fusion, and cross-modal attention.
- Design graph learning modules to capture contextual dependencies for emotion recognition in conversation.
- Explore semi-supervised and federated learning paradigms for emotion recognition to improve robustness under limited labels and privacy constraints.
- Build end-to-end PyTorch pipelines for training and evaluation, and deploy inference workflows.
- Additional research interests: active learning, handwritten mathematical expression recognition, medical federated learning, and applied computer vision for real-world applications.

Founder & Lead, AIoT Research Pioneer Club (FARPC) – FPT University — Sep 2025 – Present

- Lead a research-driven student community focused on AI/AIoT self-learning, research thinking, and academic dissemination.
- Organize weekly seminars, reading groups, and mentoring activities; support members in project execution and paper-style reporting.
- Coordinate collaboration with mentors/labs and curate learning materials for multimodal learning and applied machine learning pipelines.

AI Research Intern, Pythera AI — Jan 2025 – Apr 2025

- Worked as an AI engineer to develop an AI-based image white balance solution, improving color consistency and visual quality under diverse lighting conditions.
- Participated in model compression to reduce latency and deployment footprint while maintaining accuracy.
- Supported production deployment by exporting models to ONNX and validating inference consistency across environments.

Technologies

Languages: Python, C/C++, Java, SQL, Bash.
Speech & Signal Processing: Speech-to-text (ASR), speaker/emotion analysis, audio preprocessing, time–frequency analysis, feature extraction.
Frameworks/Libraries: PyTorch, Hugging Face Transformers, TorchAudio, NumPy, scikit-learn.
Tools: Git/GitHub, Linux, ONNX, Weights & Biases (wandb), LaTeX, Streamlit.

Grants

Multimodal Fusion in Speech Emotion Recognition — FPT University (Hanoi, Vietnam) Apr 2025 – Oct 2025

- Grant number: DHFPT/2025/10.

Selected Publications

Enhancing multimodal emotion recognition with dynamic fuzzy membership and attention fusion *(First author)* Feb 2026

Engineering Applications of Artificial Intelligence.
DOI: 10.1016/j.engappai.2025.113396. — Code: github.com/nhut-ngnn/FleSER.

- Proposes a multimodal emotion recognition approach with dynamic fuzzy membership with attention-based fusion.
- Targets robust affect modeling by adaptively weighting expressive cues across modalities.

Multimodal fusion in speech emotion recognition: A comprehensive review of methods and technologies *(First author)* Jan 2026

Engineering Applications of Artificial Intelligence.
DOI: 10.1016/j.engappai.2025.112624. — Repo: awesome-multimodal-fusion-emotion-recognition.

- Provides a structured review of multimodal fusion techniques for speech emotion recognition across audio, text, and vision.
- Summarizes trends in architectures, features, datasets, and evaluation practices to guide future research and reproducibility.

CemoBAM: Advancing Multimodal Emotion Recognition through Heterogeneous Graph Networks and Cross-Modal Attention Mechanisms *(First author)* Sep 2025

APNOMS 2025 (Kaohsiung, Taiwan).
DOI: 10.23919/apnoms67058.2025.11181320. — Code: github.com/nhut-ngnn/CemoBAM.

- Introduces a heterogeneous graph-based framework with cross-modal attention for multimodal emotion recognition.
- Models relational structure across modalities to capture interaction patterns beyond sequential fusion.

HemoGAT: Heterogeneous Multimodal Speech Emotion Recognition with Cross-Modal Transformer and Graph Attention Network *(First author)* 2025

Advances in Electrical and Electronic Engineering. To appear.

- Proposes a heterogeneous multimodal emotion recognition approach using a cross-modal Transformer and graph attention for structured fusion.
- Targets improved modeling of modality-specific cues and their interactions for emotion prediction.

ALMUS: Enhancing Active Learning for Object Detection with Metric-Based Uncertainty Sampling *(Co-author)* Sep 2025

APNOMS 2025 (Kaohsiung, Taiwan).
DOI: 10.23919/apnoms67058.2025.11181447.

- Proposed ALMUS, a metric-based uncertainty sampling approach with dynamic class-wise budget allocation to balance classification and localization.
- Validated on benchmark datasets, showing improved performance and faster convergence compared with active learning baselines.

Awards

The 27th Euréka Award for Student Research — Consolation Prize Dec 2025

- Topic: Speech Emotion Recognition with Heterogeneous Graph Construction.

Student Research Competition (Fall 2025), FPT University — Champion Prize Dec 2025

Project: Semi-supervised for Multimodal Emotion Recognition

FPT Research Festival (ResFes 2025) — Consolation Prize Aug 2025

- Topic: Speech Emotion Recognition with Heterogeneous Graph Construction.

Student Research Competition (Spring 2025), FPT University — Champion Prize Apr 2025

Project: Heterogeneous Graph Construction for Multimodal Emotion Recognition