Luong Phat Nguyen

PhD in Computer Vision | AI Research Scientist — Multimodal LLMs & Voice AI

Paris, France luongphatnguyen@gmail.com +33 6 95 32 32 97
lphatnguyen.github.io linkedin.com/in/nguyen-luong-phat github.com/lphatnguyen

PROFILE

PhD in computer vision with over 7 years of AI R&D experience, from academic research to industrial deployment. Author of four international publications, I have designed and optimised vision models, multimodal LLMs and real-time voice systems, with measurable gains in accuracy and latency. I am looking for a Research Scientist/Engineer position where I can bring this combined scientific and industrial expertise to develop novel approaches in multimodal AI and turn them into high-performance, high-impact solutions.

PROFESSIONAL EXPERIENCE

AI Research Scientist | **Rizlum AI** Mar. 2025 – Present | Paris, France

- Designed multimodal document-understanding models for insurers and health mutuals (health-care invoices, motor-accident statements): **over 95% correct extraction**, in a few seconds per document.
- Generated synthetic data and anonymised identity data (regulated sector): performance on par with real data, with no exposure of personal data.
- Built Rizlum AI's proprietary Speech-to-Text and Text-to-Speech model (**sub-200 ms latency**): low latency conversation, and no more reliance on third-party voice APIs.
- Co-developed a real-time speech-to-speech voice agent for an insurance client (architecture review, STT/TTS, evaluation): **80% of phone membership sign-ups automated**.

Technologies: Multimodal LLMs, TTS, ASR, PyTorch, Hugging Face, RAG, DeepEval, real-time inference

Postdoctoral Researcher in LLMs | **Toyota Motor Europe & LISSI** Jul. 2024 – Mar. 2025 | Paris, France

- Designed and trained vision-language multimodal LLMs (Mamba and Transformer architectures).
- Industrialised large-scale multimodal training pipelines (VQA, Conceptual Captions).
- Created an emotionally intelligent conversational agent, and published related work at **IEEE ICIP 2025**.

Technologies: Multimodal LLMs, Mamba, facial emotion recognition, PyTorch

AI Research Scientist | **Trax Retail** Dec. 2021 – Apr. 2024 | Paris, France

- Developed on-shelf product recognition models (**over 95% accuracy**, more than 500 SKUs): a photo audit taking a few seconds instead of a manual in-store check.
- Embedded real-time algorithms in Swift into an iOS SDK (CoreML, ARKit, SceneKit): on-device inference, with no network dependency and no server inference cost.
- Optimised and generalised pre-trained models without massive re-annotation: faster, cheaper roll-out to new retailers and new countries.
- Computed business indicators (*share-of-shelf*) in real time at **under 1% error and under 200 ms latency**, usable by brands during the store visit.

Technologies: PyTorch, CoreML, ARKit, Swift, OpenCV, 3D vision, augmented reality

PhD Candidate in Computer Vision | **LIFAT, University of Tours** Nov. 2018 – Dec. 2021 | Tours, France

- Carried out a thesis on spatio-temporal patterns in videos: **2 publications** (ICPR, VISAPP).
- Designed lightweight video classification architectures (R(2+1)D, R3D, LSTM, ViT) for scarce annotated data; leveraged optical flow (FlowNet2, PWC-Net) and sequential pattern mining.

Technologies: PyTorch, Keras, OpenCV, Python, C++, data mining

CORE SKILLS

AI & deep learning	Multimodal LLMs & VLMs, Text-to-Speech, Speech-to-Text, Speech-to-Speech, conversational agents, Transformers, CNN, LSTM, state-space models (Mamba), fine-tuning (LoRA / QLoRA / PEFT), model optimisation and distillation, RAG, DeepEval
Computer vision	Image and video recognition and classification, optical-flow estimation, augmented reality, 3D vision, spatio-temporal pattern extraction
Production	Real-time and low-latency inference, on-device models (CoreML), industrialisation of training pipelines, identity-data anonymisation and synthetic data generation (regulated sector)
Programming	Python, C/C++, Swift, Bash, JavaScript, Matlab, LaTeX
Frameworks & tools	PyTorch, Hugging Face, vLLM, Keras, OpenCV, CoreML, ARKit, SceneKit, Docker, Git, Linux, Google Cloud, FastAPI
Transferable	Applied research, research $\rightarrow$ product transfer, scientific publishing, university teaching, large-scale data processing

EDUCATION

PhD in Computer Vision | **University of Tours — LIFAT Laboratory** 2018 – 2021

Thesis: *Extraction and characterisation of spatio-temporal patterns in videos*.

Engineering Degree (MSc equivalent) | **INSA Centre Val de Loire** 2015 – 2018

Major: *Automated systems and industrial computing*.

PUBLICATIONS

L. P. Nguyen, H. Abdelkawy, A. Othmani. *EmoMamba: Advancing Dynamic Facial Expression Recognition with Visual and Textual Fusion*. IEEE International Conference on Image Processing (**ICIP**), 2025, pp. 1432–1437.

L. P. Nguyen, J. Mille, D. H. Li, D. Conte, N. Ragot. *Efficient Dynamic Texture Classification with Probabilistic Motifs*. International Conference on Pattern Recognition (**ICPR**), 2022, pp. 564–570.

L. P. Nguyen, J. Mille, D. Li, D. Conte, N. Ragot. *Trajectory Extraction and Deep Features for Classification of Liquid-gas Flow under the Context of Forced Oscillation*. International Conference on Computer Vision Theory and Applications (**VISAPP**), 2020, pp. 17–26.

L. P. Nguyen, M. Saleh, R. Le Bouquin Jeannès. *An Efficient Design of a Machine Learning-based Elderly Fall Detector*. International Conference on IoT Technologies for HealthCare (**HealthyIoT**), Springer, 2017, pp. 34–41.

TEACHING

- **Neural Networks — Deep Learning** (lectures and lab sessions) — 5th year, Polytech Tours, University of Tours, 2020.
- **Object-Oriented Programming — C++** (lab sessions) — 3rd year, INSA Centre Val de Loire (GSI), 2019–2020.
- **Introduction to Deep Learning** (lab sessions) — MSc 2 Big Data Management and Analytics, 2018–2019.

AWARDS

- *Rencontres du Vietnam* excellence scholarship — Odon Vallet, 2016–2018.
- Silver medal, Vietnam National Physics Olympiad, 2014.

LANGUAGES

English	Fluent (advanced)
French	Fluent (professional)
Vietnamese	Native