

Quang Minh Dinh

minh101202@gmail.com | [linkedin.com/in/quang-minh-dinh](https://www.linkedin.com/in/quang-minh-dinh) | github.com/quangminhdinh

Education

Simon Fraser University

Bachelor of Science in Computing Science

Burnaby, BC

May 2022 – June 2026

Publications

- [1] Paige Tuttösí, Mantaj Dhillon, Luna Sang, Shane Eastwood, Poorvi Bhatia, **Quang Minh Dinh**, Avni Kapoor, Yewon Jin, and Angelica Lim. “BERSting at the screams: A benchmark for distanced, emotional and shouted speech recognition”. In: *Computer Speech and Language* 95 (2026), p. 101815. ISSN: 0885-2308. DOI: <https://doi.org/10.1016/j.csl.2025.101815>.
- [2] **Quang Minh Dinh**, Hoda Rezaee Kaviani, Mehrdad Hosseinzadeh, and Yuanhao Yu. “Extended Loss: Incorporating Long Context into Training Models when using Short Audio Frames”. In: *Interspeech 2025*. 2025, pp. 778–782. DOI: [10.21437/Interspeech.2025-2177](https://doi.org/10.21437/Interspeech.2025-2177).
- [3] **Quang Minh Dinh**, Minh Khoi Ho, Anh Quan Dang, and Hung Phong Tran. “TrafficVLM: A Controllable Visual Language Model for Traffic Video Captioning”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. June 2024, pp. 7134–7143.

Preprint

- [4] **Quang Minh Dinh***, Stella Lin*, Gemmin Sugiura*, Bitu Azari, Yasaman Etesam, Chuxuan Zhang, and Angelica Lim. “Non-Verbal Human-Robot Interaction with Reachy Mini: A Real-Time Multimodal System and Turing Test Evaluation”. 2026. URL: <https://quangminhdinh.github.io/nonverbal-reachy>.

Research Experience

Rosie Lab, Simon Fraser University

Student Researcher (Supervisor: Angelica Lim)

Burnaby, BC

Jan. 2026 – Present

- Built a real-time human-to-robot interaction pipeline with Mediapipe hand, pose, and facial tracking, LLM-based reasoning, and action boundary detection and stabilization on pose landmarks.
- Designed a facial video frames matching system using OpenFace 3.0 action unit and emotion detection, utilizing temporal pooling and dynamic time wrapping for aggregation.

Student Researcher (Supervisor: Angelica Lim)

Nov. 2024 – Jan. 2025

- Fine-tuned several variations of DAWN and Odyssey WavLM on 4523 samples of the BERSt dataset for speech emotion recognition and analyzed the results.
- Developed different SVM models using the extracted valence, dominance, and arousal data extracted from SOTA models for speech emotion recognition.

APEX Lab, Simon Fraser University

Student Researcher (Supervisor: Ke Li)

Burnaby, BC

May 2025 – Aug. 2025

- Extended Implicit Maximum Likelihood Estimation (IMLE) to the text-conditional setting, designing and benchmarking four conditioning architectures including FiLM-style affine modulation, per-block residual injection, StyleGan-based conditioning, and mapping-network concatenation.
- Evaluated conditioning strategies on more than 8000 samples of the Oxford 102 Flowers and 10000 samples of the CelebA Faces Captioned datasets to identify optimal text-to-image generation approaches.

Professional Experience

Huawei Technologies Canada Co., Ltd.

Assistant Engineer Intern

Markham, ON

May 2023 – Apr. 2024

- Implemented the end-to-end training and evaluation pipelines for CRUSE, Align-CRUSE, and different Conformer and DeepVQE variations for Acoustic Echo Cancellation (AEC).
- Designed and implemented different audio synthesizer pipelines to generate online training data using 350K speech audio samples, 62K noise samples, and 60K room impulse responses from the Deep Noise Suppression Challenge.

- Proposed a new loss calculation method that boosts the training performance on short audio frames for real-time speech enhancement applications, resulting in an average increment of 0.1-0.57 double-talk AECMOS echo and 5.69-24.18 far-end single-talk ERLE for all experiments on the AEC Challenge 2023 aligned blind set.

Mount2 Speak *(early-stage AI startup)*

Remote

Machine Learning Engineer, Part-time

Dec. 2022 – Mar. 2023

- Fine-tuned and benchmarked different vision transformer models on 28,709 samples of the FER2013 dataset for face emotion detection.
- Implemented a real-time face emotion detection solution using MediaPipe Face Mesh and angular encoding.
- Optimized the inference model using ONNX and TensorRT, and deployed on Banana.dev serverless infrastructure.

Front-end Developer, Part-time

Sep. 2022 – Dec. 2022

- Developed the web front-end for a platform that supports high-school students in developing public speaking skills using automated speech recognition using Node.js, React, TypeScript, Next.js and MongoDB.

YITEC Research and Technology Development Co. Ltd

Hanoi, VN

Machine Learning Engineer Intern

Oct. 2020 – June 2021

- Developed a real-time hand tracking and AR ring try-on mobile application using MediaPipe, Unity, Android Studio and AR Foundation.
- Utilized Google ARCore to mimic frames' lighting cues and render the virtual ring under the same conditions.
- Experimented with different video stabilization and vibration denoising algorithms to increase the frame rate from 5 FPS to 30 FPS.

Personal Projects

TrafficVLM: A Controllable Visual Language Model for Traffic Video Captioning

- Developed a novel visual language model for temporal localization and dense traffic video captioning, utilizing two Vid2Seq-based visual encoders to model traffic video events at different levels of analysis and a T5-Base text decoder to generate long fine-grained descriptions for both vehicle and pedestrian at different traffic phases.
- Devised a multi-task fine-tuning paradigm, allowing the model to effectively learn the alignments between the video and textual features for all traffic event phases.
- The solution achieves the final average score of 32.3006, ranking 3rd in Track 2 of the AI City Challenge 2024.

MirrorBrain | *Next.js, TypeScript, Tailwind, BlockNote, Python, FastAPI, Qdrant, Convex* Oct. 2023 – Feb. 2024

- Led a team of 5 people to develop a Zettelkasten-style note-taking web application augmented with LLM in auto-generating text connections and contexts for the web-of-thoughts.
- Designed the system architecture, wireframe UX prototype, database schema, REST API endpoints, and managed the project using the Kanban method with Trello.
- Reverse engineered BlockNote to overwrite the side menu UI and trigger command.
- Served the LLM batch-processing services on Railway.

Poster Talks

Real-Time Non-Verbal Human Robot Interaction with Reachy Mini

- IEEE Vancouver Section Annual General Meeting (AGM) 2026. Mar. 28, 2026
- SFU Undergraduate Research Symposium (UGRS) 2026. Apr. 1, 2026

Awards & Honors

Dean's Honour Roll, Simon Fraser University (Fall 2024, Spring 2025, Fall 2025, Spring 2026).

Undergraduate Student Research Award (USRA) Nominee, Simon Fraser University (2025).

Finalist of Social Business Creation, HEC Montréal (2022).

Impactful Social Business Idea Award - Social Business Creation, HEC Montréal (2022).

International Summit Scholarship, Simon Fraser University (2022).

Services

Volunteer, AI/CRV 2026.

Reviewer, ICCVW AI City Challenge 2025.

Reviewer, CVPRW AI City Challenge 2024.