TRUONG-BINH DUONG

+84 936 212 865 | Montréal, QC, Canada | duongtruongbinh2003@gmail.com
LinkedIn | GitHub | Google Scholar

PROFILE

M.Sc. student in Telecommunications at the **Institut national de la recherche scientifique (INRS)**, supervised by Prof. Long Bao Le. My current research interests include world models, vision-language-action (VLA) models, embodied AI, and robot learning, with broader interests in multimodal learning and robust AI. My previous research focused on robust visual question answering, vision-language models, benchmark construction, and multimodal evaluation.

EDUCATION

M.Sc. in Telecommunications, Institut national de la recherche scientifique (INRS) Sep 2026 – Present
Thesis-based Master's Program

- Supervisor: Prof. Long Bao Le
- Research interests: world models, vision-language-action models, embodied AI, and learning-based robotic decision making.

B.Sc. in Information Technology, University of Science - VNUHCM 2021 – 2025

- GPA: 3.83/4.0 | Dean's List (5 semesters)
- Graduation Thesis: *Counterfactual Reasoning for Robust Visual Question Answering*

SELECTED PUBLICATIONS

Truong-Binh Duong, Thanh-Ngan Tran, Ngoc-Thao Nguyen, Bac Le. "Counterfactual Reasoning for Robust Visual Question Answering." *Accepted at the 30th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES 2026); to appear in Procedia Computer Science (Elsevier).* GitHub

- Proposed a three-stage counterfactual contrastive-learning framework with Batch-Contrastive, Answer-Contrastive, and Gradient-Discrepancy losses to improve causal visual grounding under language-prior shifts.
- Achieved **61.64%** accuracy on VQA-CP v2 and **62.80%** on VQA v2, yielding a **1.16 %** OOD–ID generalization gap.

Truong-Binh Duong, Hoang-Minh Tran, Binh-Nam Le-Nguyen, Dinh-Thang Duong. "An Automated Pipeline for Constructing a Vietnamese VQA-NLE Dataset." *In Proc. ICISN 2025 (Springer LNNS, Vol. 1596).* Springer | GitHub

- Designed and implemented an automated pipeline leveraging multiple LLMs as evaluators to build ViVQA-X, the first Vietnamese visual question answering (VQA) dataset with natural language explanations.

Tue-Thu Van-Dinh*, Hoang-Duy Tran*, **Truong-Binh Duong**, et al. (*equal contribution). "ViInfographicVQA: A Benchmark for Single and Multi-Image Visual Question Answering on Vietnamese Infographics." *In Proc. AAAI Workshop on AI for Scientific Research*, 2026. arXiv | GitHub

- Spearheaded data acquisition for 6,000+ infographics and benchmarked state-of-the-art VLMs to establish the first Vietnamese benchmark for single-image and multi-image visual reasoning.

Yen-Linh Vu*, Dinh-Thang Duong*, **Truong-Binh Duong**, et al. "Describe Anything Model for Visual Question Answering on Text-rich Images." *In Proc. VisionDocs Workshop @ ICCV*, 2025. arXiv | GitHub

- Led the comprehensive evaluation and experimental validation of multiple VLMs across six benchmark datasets for text-rich VQA.

RESEARCH EXPERIENCE

Counterfactual Reasoning for Robust Visual Question Answering Jan 2025 – Aug 2025
B.Sc. Graduation Thesis | Supervisor Prof. Hoai Bac Le GitHub

- Developed a robust VQA training framework combining counterfactual sample synthesis, Batch-Contrastive learning, Answer-Contrastive regularization, and Gradient-Discrepancy supervision for causal visual grounding. Resulted in a paper accepted at KES 2026

Research & Teaching Assistant Aug 2024 – Present
AI VIETNAM *Remote*

- **Research**: Conduct research on VLMs and LLM reasoning, with a focus on automated data pipelines, benchmark construction, and empirical evaluation of multimodal systems.
- **Teaching Assistant**: Develop curriculum, materials (Course Materials), and comprehensive exams for the AIO program, covering topics from foundational mathematics to deep learning and LLM/VLM.

TECHNICAL SKILLS

Technical	Python, PyTorch, Transformers, Hugging Face, OpenCV, C/C++, SQL, Scikit-learn, Docker
Languages	Vietnamese (Native), English (IELTS Academic: 7.0)