

NGOC DUNG HUYNH

PhD Researcher | ML Engineer | Multi-modal AI Specialist

Deakin University (Australia) | Research Consultant, TII (UAE)

dunghuynhparker@gmail.com

+61 412 697 643

Melbourne, VIC, Australia

dunghuynhandy

dunghuynhandy

dunghuynh

Professional Summary

PhD researcher and ML engineer with **4+ years** of experience architecting large-scale multi-modal AI systems, production data pipelines, and vision-language models. Delivered end-to-end solutions processing **3M+ PDFs**, curating internet-scale datasets, and deploying VLM/LLM systems on GCP and AWS. **First-author publications at ICCV 2025 and CVPR 2026**. Deep expertise in PyTorch, HuggingFace Transformers, OCR pipelines, and distributed ML infrastructure. Ranked **Top 9 worldwide** (Toloka VQA Challenge) and **Top 7 globally** (COVID Detection Challenge).

Core Competencies

ML / AI	Vision-Language Models (VLMs), LLMs, Visual Question Answering, Multi-modal Reasoning, Speech-Vision-Language
Frameworks	PyTorch, TensorFlow, Keras, HuggingFace Transformers, Weights & Biases
Data Engineering	OCR Pipelines, ETL, Web Crawling, Deduplication, LLM Filtering, SFT Data Generation, Agent-Based Pipelines
Infrastructure	AWS, GCP, Docker, Linux, Slurm, Flask, React.js, Elasticsearch
Programming	Python, JavaScript, SQL, R, C++
Specialties	Annotation Systems, STEM-VQA, Distributed Training, Benchmark Evaluation

Professional Experience

Technology Innovation Institute (TII)

Engineering Consultant — Multi-modal AI & Data

Abu Dhabi, UAE (Remote) • Jan 2025 – Present

- **Large-Scale Document & Web Data Processing**
 - Architected and deployed production-grade ETL pipelines to crawl, deduplicate, and normalize internet-scale multi-modal datasets, directly supporting model training workflows for Falcon-H
 - Processed **3M+ PDFs** via OCR, layout parsing, and CV-based structured text extraction; engineered multi-stage cleaning, normalization, and deduplication to maximize downstream dataset quality
 - Synthesized large-scale **Supervised Fine-Tuning (SFT)** datasets using GPT-4, Gemini, Claude, and Qwen, directly accelerating model alignment timelines for Falcon-H production releases
- **Multi-modal & Agent-Generated Dataset Engineering**
 - Unified data from **10+ agent platforms** into multi-modal corpora; built content filtering, metadata normalization, and quality-scoring pipelines to ensure training-ready data at scale
 - Engineered a **React-based annotation platform** supporting segmentation and bounding-box labeling, significantly increasing annotation throughput for multi-modal model teams
 - Trained large-scale VLMs on **GCP and AWS**, optimizing GPU utilization and job scheduling across distributed Slurm clusters
- **VQA & STEM-VQA Systems**
 - Built end-to-end VQA training and evaluation pipelines for scientific domains including STEM, charts, equations, and scientific plots, enabling systematic benchmarking of emerging VLMs
 - Designed topic-aware sampling and data filtering strategies to curate high-quality STEM-VQA alignment datasets, improving model performance on held-out evaluation benchmarks

Technology Innovation Institute (TII)

Research Intern — Multi-modal AI

Abu Dhabi, UAE • Apr 2024 – Jan 2025

- Led research on multi-modal reasoning across speech, vision, and language modalities, contributing to published findings at ICCV 2025
- Developed and integrated ASR, VQA, OCR, and LLM inference components into unified end-to-end pipelines, advancing prototype-to-research-ready system quality
- Co-authored ICCV 2025 paper and contributed to multiple arXiv publications on VLM evaluation and multi-modal reasoning

Deakin University
Research Assistant — Visual Question Answering

Melbourne, Australia • Mar 2022 – Oct 2022

- Ranked **Top 9 worldwide** in the Toloka VQA Challenge (WSDM Cup 2023), competing against thousands of global participants
- Achieved **Top 7 globally** in the COVID Detection Challenge, applying computer vision to real-world medical imaging
- Conceived and launched a university-wide AI competition at Deakin University, growing the campus ML community and driving student engagement across faculties

Stealth Startup
Software Engineer (Part-time)

Singapore (Remote) • Aug 2020 – Nov 2021

- Designed and delivered a full-stack, multi-user annotation platform (Flask + React + Elasticsearch) with RESTful APIs, optimizing annotation workflows and throughput at scale

Education

Deakin University
PhD in Computer Science

Melbourne, Australia • Oct 2022 – Present

Thesis: *Designing Scalable and Interpretable Vision–Language–Speech Systems for Generalised Multi-modal Reasoning*

Deakin University
MSc in Data Science (GPA: 86%)

Melbourne, Australia • Mar 2020 – Mar 2022

Thesis: *Speech-to-CDQL — Context Definition and Query Language from Natural Language for Smart Home*

University of Education Hue
BSc in Mathematics

Vietnam • Sep 2014 – Sep 2018

Selected Publications

- **Brigitta T., Dahou Y., Huynh N. D., et al.** VisRes Bench: Evaluating Visual Reasoning Capabilities of VLMs. *CVPR 2026*.
- **Huynh N. D., et al.** SVLA: A Unified Speech-Vision-Language Assistant. *arXiv:2503.24164, 2025*.
- **Dahou Y., Huynh N. D., et al.** Vision-Language Models Can't See the Obvious. *ICCV 2025*.
- **Huynh N. D., et al.** Visual Question Answering: A Survey. *arXiv:2501.03939, 2025*.
- **Huynh N. D., et al.** Improving VQA Through Topic-Aware Selection Layer. *SSRN:5385867, 2024*.
- **Huynh N. D., et al.** SimpsonsVQA. *arXiv:2410.22648, 2024*.
- **Huynh N. D., et al.** Jarvis: A Voice-based Context-as-a-Service Tool. *IEEE MDM, 2023*.
- **Zuo J., et al.** [Contributing Author] Falcon-H1: Hybrid-Head Language Models. *arXiv:2507.22448, 2025*.

Honours & Awards

- **Best Demo Paper** IEEE MDM, 2022
- **Top 9 Worldwide — Toloka VQA Challenge** WSDM Cup, 2023
- **Top 7 Globally — COVID Detection Challenge** 2022
- **1st Prize — Simpsons Character Classification Competition** Deakin University, 2021
- **MSc Scholarship (2 years)** Deakin University, 2020–2021