Quy-Anh Dang

Hanoi, Vietnam
dangquyanh150101@gmail.com
GitHub: QuyAnh2005 | LinkedIn: Quy-Anh Dang | Google Scholar Homepage

Summary

Machine Learning Researcher with 2+ years of experience developing production-ready AI solutions, specializing in Large Language Models and MLOps systems. Demonstrated research excellence through Top 3 ranking in NeurIPS LLM 2023 Competition and multiple publications. Strong mathematical foundation combined with hands-on expertise in building end-to-end ML pipelines, retrieval-augmented generation systems, and scalable infrastructure. Focused on advancing research in LLM safety, trustworthy AI, efficient model training, and speech processing.

Research Interests

My research centers on three core areas: (1) **AI Safety and Alignment** – developing trustworthy AI systems through red teaming, adversarial robustness, and model alignment techniques; (2) **Efficient Machine Learning** – advancing parameter-efficient fine-tuning methods (QLoRA, LoRA), model merging strategies, and data-centric approaches to optimize training under computational constraints; and (3) **Speech Processing** – exploring text-to-speech synthesis and speech recognition systems. I am particularly interested in the intersection of these areas, investigating how to build safe, efficient, and practical AI systems that can be deployed responsibly in real-world applications.

Education

Master of Science in Data Science 2023 – 2025
VNU University of Science *Hanoi, Vietnam*

GPA: 3.95 / 4.0 **Class Rank:** 1st **Department Rank:** 1st

Bachelor of Science in Applied Mathematics and Computer Science 2019 – 2023
VNU University of Science *Hanoi, Vietnam*

GPA: 3.71 / 4.0 **Class Rank:** 1st **Department Rank:** 2nd out of 350+ students

Work Experience

Machine Learning Researcher Jan 2024 – Present
Knovel Engineering *Full-time Remote – Singapore*

- **DeepAssure Platform Development:** Serve as core research team member developing DeepAssure, an AI governance and trustworthy AI platform. Implement core algorithms, and optimize platform infrastructure to ensure responsible AI deployment and build confidence in AI decision-making systems.
- **Singapore Police AI Chatbot Project:** As a member of AI Verify Foundation, lead data curation and dataset synthesis for the Singapore Police Force's AI chatbot initiative. Conduct comprehensive performance testing and adversarial robustness evaluation against jailbreak attacks to ensure system safety. Contribute to safety assurance for public deployment of the AI assistant for police report lodging. Published assurance report demonstrating trustworthy AI practices.

- **RySense AI Infrastructure:** Design and implement production-ready RAG-based AI infrastructure for RySense, Singapore's leading social research organization. Enhance their intelligent survey platform with advanced natural language understanding and automated insight generation capabilities.
- **AI Safety & Alignment Research:** Conduct research on LLM safety, red teaming methodologies, and alignment techniques. Focus on adversarial robustness testing, comprehensive model evaluation frameworks, and development of safety protocols for trustworthy AI systems.
- **AI Application Development:** Develop proof-of-concept applications including document processing pipelines integrating frontier models with LlamaIndex/Langchain. Train and deploy optimized open-sourced LLMs for structured data querying (CSV/Excel). Build speech processing models for text-to-speech and speech recognition. Maintain MLOps evaluation pipelines using Weights & Biases, MLflow, and TrueLens.

Full-Stack Data Scientist Jun 2023 – Sep 2024
Wingsmob Studio *Hanoi, Vietnam*

- Designed and implemented automated ETL pipelines to extract data from Adjust API, store in Amazon S3, and load into BigQuery via Google Cloud Functions, processing 10M+ daily events.
- Developed dynamic, real-time dashboards using Power BI to monitor user engagement metrics, resulting in 50% reduction in user dropout rates through data-driven UX optimizations.
- Built and deployed predictive ROAS (Return on Ad Spend) estimation model achieving <5% error rate, enabling optimized marketing budget allocation and improving campaign ROI by 30%.
- Collaborated with product and marketing teams to translate business requirements into analytical solutions and actionable insights.

Publications

Quy-Anh Dang, Chris Ngo, Truong-Son Hy. "RainbowPlus: Enhancing Adversarial Prompt Generation via Evolutionary Quality-Diversity Search." Under Review, *Main Conference @ EMNLP*, 2026.

Quy-Anh Dang, Chris Ngo. "Selective Steering: Norm-Preserving Control Through Discriminative Layer Selection." *Findings @ ACL*, 2026.

Quy-Anh Dang, Chris Ngo, Truong-Son Hy. "RedBench: A Universal Dataset for Comprehensive Red Teaming of Large Language Models." *Principled Design for Trustworthy AI - Interpretability, Robustness, and Safety across Modalities @ ICLR*, 2026.

Quy-Anh Dang, Chris Ngo. "Polyglot-Lion: Efficient Multilingual ASR for Singapore via Balanced Fine-Tuning of Qwen3-ASR." *Multilinguality in the Era of Large Language Models @ ACL*, 2026.

Quy-Anh Dang, Chris Ngo. "Reinforcement Learning for Reasoning in Small LLMs: What Works and What Doesn't." *Logical and Symbolic Reasoning in Language Models @ AAAI*, 2026.

Quy-Anh Dang, Chris Ngo. "MoD: A Distribution-Based Approach for Merging Large Language Models." *LLM Merging Competition @ NeurIPS*, 2024.

Teaching Experience

Teaching Assistant 2022 – 2024

- **The University of Alabama at Birmingham:** CS660/CS760 Artificial Intelligence (Fall 2024)
- **DataScience.Kan:** Generative AI (2024, K2 & K3), Machine Learning Operations (2023), Machine Learning Hands-on (2022), Math for Machine Learning (2022)
- **VNU University of Science:** Data Structures and Algorithms (2022)
- Assisted in curriculum development, conducted lab sessions, mentored students on projects, and evaluated assignments for 30+ students per course.

Technical Skills

Programming & Frameworks: Python (Pandas, NumPy, Scikit-learn, PyTorch, TensorFlow, PySpark, Keras), SQL

Machine Learning: Deep Learning (MLP, CNN, LSTM, RNN, Transformers), Classical ML (Linear/Logistic Regression, SVM, Decision Tree, Random Forest, Gradient Boosting), Feature Engineering

AI & NLP: Large Language Models (GPT-4, LLaMA), Fine-tuning (QLoRA, LoRA), RAG Systems, LlamaIndex, Prompt Engineering

MLOps & Tools: Weights & Biases, MLflow, TrueLens, Docker, FastAPI, Git, Linux

Data Engineering: Apache Spark, ETL Pipelines, BigQuery, Cloud Functions, Data Warehousing

Data Visualization: Google Data Studio, Power BI, Tableau, Plotly

Mathematics: Strong foundation in Statistics, Linear Algebra, Optimization, Probability Theory

Honors & Awards

- **Winter Mathematics Camp 2025** – National University of Singapore Scholarship Recipient
- **NeurIPS LLM Efficiency Challenge 2023** – Top 3 (4090 Track) and Top 8 (A100 Track)
- **University Scientific Research Conference 2023** – First Prize
- **International Data Analysis Olympiad 2022** – Top 6 Finalist (HSE University & Yandex)
- **Outstanding Academic Achievement Award** – 2020-2021, 2021-2022, 2022-2023 (Three consecutive years)

Certifications

- **Coursera:** Machine Learning Specialization, Deep Learning Specialization, GANs Specialization, Natural Language Processing Specialization, Machine Learning Engineering for Production (MLOps) Specialization
- **Codecademy:** Data Science Career Path (2021)
- **Facebook Developer Circles:** Data Science Track (2020)