

TUE VU

vuminhtue@gmail.com | 864-624-3868 | Dallas, Texas

LinkedIn: [linkedin.com/in/tue-vu](https://www.linkedin.com/in/tue-vu) | GitHub: github.com/vuminhtue | HuggingFace: huggingface.co/vuminhtue

Google Scholar: scholar.google.com/citations?user=IS9y7dEAAAAJ | NVIDIA GTC 2022 Speaker

PROFESSIONAL SUMMARY

Machine Learning Engineer and AI Research Scientist with 10+ years of experience designing, training, and deploying deep learning models on GPU-accelerated infrastructure. Expert in end-to-end ML pipelines spanning computer vision, NLP, and generative AI, with hands-on production experience on NVIDIA SuperPOD, A100, V100, and Jetson platforms. NVIDIA DLI Ambassador with demonstrated proficiency in CUDA, RAPIDS, multi-GPU distributed training, and large-scale model serving. PhD in Computational Informatics with 80+ workshops delivered and publications cited 3000+ times.

TECHNICAL SKILLS

GPU Computing & HPC:

CUDA, NVIDIA RAPIDS (cuDF, cuML, cuGraph, cuXfilter), TensorRT, Horovod, NCCL, MPI, OpenMP, NVIDIA Jetson Nano, NVIDIA SuperPOD, SLURM/PBS job scheduling, Spack, Lmod, Open OnDemand, Docker, Singularity

Deep Learning Frameworks:

PyTorch, TensorFlow/Keras, Hugging Face Transformers, NVIDIA NeMo, Stable Diffusion (Automatic1111, ComfyUI, InvokeAI), LangChain, Ollama

ML & AI Techniques:

CNNs (ResNet, VGG, EfficientNet), Transformers/Attention Mechanisms, LLMs, RAG Pipelines, Transfer Learning, LSTM, Autoencoders, GANs, Diffusion Models, XGBoost, SVM, KNN, LASSO, PCA, DBSCAN, Ensemble Methods, Feature Engineering, Anomaly Detection, Recommender Systems, Predictive Maintenance

Programming & Data:

Python (expert), R, SQL, C/C++ (familiar), MATLAB, NumPy, Pandas, scikit-learn, Dask, PySpark, Beautiful Soup

MLOps & Cloud:

Docker, Singularity, REST APIs, Streamlit (web deployment), Git/GitHub, AWS, Azure, Databricks

Visualization & BI:

Tableau, Power BI, Rapidminer, Matplotlib, ggplot2, D3.js, Plotly

PROFESSIONAL EXPERIENCE

Southern Methodist University - Dallas, TX

Nov 2021 - Present

AI & ML Research Scientist | Lead Data Scientist - Office of Information Technology

Key Roles: SMU Local AI Expert | Generative AI Ambassador | NVIDIA DLI Ambassador | HPC Facilitator (ManeFrame II, III, NVIDIA SuperPOD)

GPU-Accelerated Computing & Multi-GPU Optimization

- Benchmarked and optimized multi-GPU distributed training across NVIDIA P100, V100, and A100 architectures using Horovod and PyTorch DDP for computer vision workloads, achieving near-linear scaling efficiency
- Deployed billion-parameter LLMs on NVIDIA SuperPOD GPU cluster for high-throughput inference, orchestrating model serving through Ollama backend
- Built GPU-accelerated big data pipeline using NVIDIA RAPIDS (cuDF, cuML, cuGraph, cuXfilter) and Dask, processing terabytes of cellphone-location records for traffic behavior analysis
- Accelerated unsupervised clustering on massive geospatial datasets using GPU-accelerated DBSCAN via RAPIDS cuML, with visualization through cuXfilter

Generative AI & LLM Applications

- Designed and deployed RAG pipelines using LangChain + Ollama with very large LLMs (100B+ parameters) on SuperPOD, ingesting NASDAQ quarterly filings for financial insight extraction

- Built production RAG chatbot with Python Streamlit frontend and LangChain orchestration, enabling faculty to query private journal article collections and DOD materials
- Developed multimodal RAG pipeline incorporating OpenAI Whisper ASR for video-to-text transcription, enabling on-demand Q&A from video lecture content
- Engineered on-premise RAG system that condenses thousands of detective/police reports into actionable summaries and builds predictive models for report authenticity detection
- Created image-centric multimodal chatbot letting faculty interact with private visual collections using billion-parameter vision-language models on NVIDIA SuperPOD GPUs
- Trained faculty to deploy Stable Diffusion WebUIs (Automatic1111, ComfyUI, Fooocus, InvokeAI) on SuperPOD NVIDIA GPUs for text-to-image generation

Computer Vision & Transfer Learning

- Conducted transfer learning with TensorFlow on pretrained CNN architectures (ResNet, VGG, EfficientNet) for object recognition, optimizing training across multiple NVIDIA GPU architectures; deployed inference on NVIDIA Jetson Nano for real-time edge demonstration
- Developed computer vision pipeline for ancient artifact identification using deep learning

NLP & Transformer Models

- Built Transformer-based models (BERT, GPT variants) to predict Big Five personality traits (NEO-PI) from interview transcripts and questionnaires
- Executed large-scale web scraping and NLP data-mining pipeline using Python, gathering structured data on parliamentary members across 15 African nations

Anomaly Detection & Predictive Analytics

- Applied Autoencoder and GAN architectures for anomaly detection on HPC network packet data, identifying malicious traffic patterns
- Implemented LSTM-based predictive maintenance model for proactive HPC infrastructure monitoring
- Built classification models to predict student academic enrollment success for CIO office strategic planning

Clemson University - Clemson, SC

Aug 2015 - Nov 2021

Senior HPC Facilitator | Senior Data Scientist | Research Assistant Professor

Key Roles: Senior HPC Facilitator (Palmetto - Top500) | Senior Data Scientist | NVIDIA DLI Ambassador

- Led ML development on NIH-funded project using XGBoost to determine Post Mortem Interval (PMI) from geospatial and climatological features, delivering production-grade predictive model for forensic science
- Developed statistical analysis pipeline in R for multi-university psychological study (7 universities); resulting PLOS ONE publication cited 1,200+ times
- Applied LASSO and PCA for dimensionality reduction on large-scale climate variables, then built SVM-based supervised learning model to predict streamflow patterns
- Built stochastic weather generator using K-Nearest Neighbor approach, synthesizing meteorological variables influenced by ENSO patterns
- Administered Clemson's Palmetto HPC cluster (Top500 ranked), supporting hundreds of researchers across computational science disciplines

National University of Singapore - Singapore

Jul 2007 - Jul 2015

Computational Research Engineer & Research Fellow - Tropical Marine Science Institute

- Applied Artificial Neural Networks (ANN) to model relationships between ocean floor rupture scenarios and tsunami propagation parameters
- Used PCA for dimensionality reduction on 250 climate grid variables, then trained ANN to construct climate-to-precipitation downscaling models
- Performed K-means clustering for unsupervised climate zone classification across Southeast Asia
- Processed and analyzed large-scale satellite data (HDF/NetCDF) and managed data pipelines with missing data imputation (bag impute, KNN impute, Box-Cox transformation)

EDUCATION

Southern Methodist University, USA - 2025

Master of Science in Data Science (MSDS)

Specialization: Machine Learning & Generative AI

National University of Singapore, Singapore - 2013

Doctor of Philosophy (PhD) - Computational Hydro-Informatics

Stanford University, USA & Nanyang Technological University, Singapore - 2007

Master of Science (MSc) - Full Scholarship

Singapore-Stanford Partnership Program - Environmental Informatics

HONORS & AWARDS

- 2025: J. James R. Croes Medal - ASCE Best Paper Award
- 2025: Best Capstone Presentation - SMU Master of Science in Data Science
- 2025: Marquis Who's Who - Honored Listee
- 2013: Best Student Poster Award - Asia Oceania Geoscience Society
- 2006: Full Scholarship - Singapore-Stanford Partnership (Master's degree)

NVIDIA & DEEP LEARNING CERTIFICATIONS

NVIDIA DLI Ambassador	2022
NVIDIA DLI - Generative AI with Diffusion Model	2024
NVIDIA DLI - Applications of AI for Predictive Maintenance	2024
NVIDIA DLI - Application of AI for Anomaly Detection	2022
NVIDIA DLI - Building Transformer-Based NLP Applications	2022
NVIDIA DLI - Fundamentals of Accelerated Data Science with RAPIDS	2022
NVIDIA DLI - Fundamentals of Deep Learning	2022
NVIDIA DLI - Fundamentals of Deep Learning for Multi-GPUs	2022
DeepLearning.AI - NLP with Attention Models (Coursera)	2023
DeepLearning.AI - NLP with Classification and Vector Spaces (Coursera)	2023
DeepLearning.AI - NLP with Sequence Models (Coursera)	2023
IBM - Building Deep Learning Models with TensorFlow (Coursera)	2021
IBM - Intro to Deep Learning & Neural Networks with Keras (Coursera)	2021
IBM - Data Engineering and ML using Spark (Coursera)	2023
Databricks - Apache Spark SQL for Data Analysts	2023
Databricks - Data Science Fundamentals for Data Analysts	2023
Databricks - Applied Data Science for Data Analysts	2023

WORKSHOPS & COMMUNITY IMPACT

- Delivered 80+ workshops to hundreds of students (undergrad to postdoc) on Research Computing, Machine Learning, Deep Learning, and Generative AI at SMU and Clemson University
- Created and conducted Generative AI workshops: Prompt Engineering with ChatGPT, Stable Diffusion GUI deployment on NVIDIA SuperPOD
- As NVIDIA DLI Ambassador, delivered official NVIDIA Deep Learning Institute workshops: Fundamental Deep Learning, Multi-GPU Training, Accelerated Data Science with RAPIDS, Transformer-based NLP with NeMo, AI for Anomaly Detection, Recommender Systems, Predictive Maintenance