NGUYEN VAN BINH

(+65) 83073859 | nguyen.binh@u.nus.edu | linkedin.com/in/nvbinh/ | github.com/nvbinh15 | nvbinh.com | medium.com/@nvbinh

SKILLS

Programming Languages	Python, C/C++, Java, Rust, RISC Assembly
AI/ML Domains and Tools	Computer Vision, NLP, Graph ML, MLOps	Scikit-Learn, PyTorch, TensorFlow, MLflow
Database/Data Processing	RDBMS, PostgreSQL, Hadoop MapReduce, Spark, DVC
Web Development	FastAPI, Django, Django REST Framework, HTML/CSS/JavaScript
Others	Docker, Cloud Platforms (Amazon Web Services, Azure), Computer Networks, Airflow, Linux

EDUCATION

National University of Singapore   **Aug 2020 – Present (Expected: May 2024)**

Bachelor of Engineering, Computer Engineering, Minors in Management and Data Engineering

- Grade Point Average (GPA): 4.91/5.0 – Dean's List in all eligible semesters
- NUS Overseas Colleges Program, NOC Paris (2023): participated in entrepreneurship program at Université PSL
- Placed on the Honour List of Student Tutors for Excellence in Teaching as a teaching assistant for courses including Data Structure & Algorithms (Master-level) and Database Systems

EXPERIENCE

Software Engineer, Machine Learning Intern, Vebits (Singapore)   **Jan 2024 – Present**

- Train and quantize ResNet-50 and YOLOv7 models with ByteTrack for image classification, object detection and tracking
- Build real-time video streaming system (1080p resolution at 30fps) integrating machine learning inference on Rockchip and Qualcomm SoC with hardware-accelerated codecs using GStreamer, multiprocessing, and pipelining in Python
- Create Python binding for C++ Snapdragon kernel libraries, streamlining development cycle for on-chip machine learning systems, bringing down inference time by 25 times compared to pure Python implementation
- Deliver tools for camera calibration, image orthogonalization, image stitching, distance, and depth estimations with OpenCV, speeding up internal image processing workflow by 55%

Data Science Intern, Science Feedback (Paris, France)   **Feb 2023 – Aug 2023**

- Developed web domain credibility ranking models with graph neural networks adopting PyTorch and NetworkX, set up end-to-end ML pipeline, deployed models in production (making prediction for > 1 million domains) using FastAPI
- Established data infrastructure (data warehouse, computation, orchestration) on Azure Cloud Services, enhancing data security, quality, and integrity, improving efficiency of data and software teams, and fastening ML life cycle by 40%
- Increased efficiency of social media data scraping process by 30% through customization of open-source tools

Machine Learning Engineer Intern, Neuron Mobility (Singapore)   **May 2022 – Aug 2022**

- Trained and deployed ML models for scooter parking image classification on AWS SageMaker achieving 95% accuracy and 0.98 precision score, detecting bad parking behaviours, enhancing user experience, and saving US$200,000 annually
- Proposed and coded backend for an in-house annotation platform utilizing Django REST Framework and PostgreSQL, streamlining collaboration between ML and data annotation teams, leading to time savings of 5+ hours per week

PROJECTS

Quora Duplicate Question Detection and Generation (source code | report)

- Led team of 6, built duplicate question identification models including RNN, GRU, LSTM, fastText and word2vec embeddings, and fine-tuned BERT family LLMs culminating in 90.1% test accuracy rate and 0.87 F1-score
- Constructed generative language models harnessing byte-pair encoding, n-grams, MLP, and decoder-only transformers accomplishing test 76.8 perplexity

Laser Tag (demo | source code): Augmented Reality shooting game

- Created an ML model classifying player's action attaining 96.14% test accuracy based on sensor data running on FPGA
- Designed architecture, optimized inter-component communication overhead through 20% data packet size reduction

AWARDS & HACKATHON

- Bronze Medal, European Physics Olympiad 2019: represented Vietnam National team in a continent-wide competition
- Champions, National AWS DeepRacer League 2019: formulated reward function and optimized the best reinforcement learning model loaded to autonomous car completing a racing lap in 9.625s

EXTRA-CURRICULAR ACTIVITIES

- Founder & Technical Writer, Binary Dissect – weekly newsletter: help hundreds of engineers level up by sharing intermediate and advanced programming and computer science concepts and techniques
- Community management, GoodHub SEA: connect with NGOs and social enterprises in Southeast Asia to deliver technical solution and consultation; organize quarterly community meetups having over 25 attendees