NGUYEN VAN BINH

(+65) 8307 3859 • nguyen.binh@u.nus.edu • linkedin.com/in/nvbinh/ • github.com/nvbinh15 • nvbinh.com • medium.com/@nvbinh

FULL-STACK MACHINE LEARNING ENGINEER

Senior Computer Engineering undergraduate at NUS specializing in full data science and engineering lifecycle. A knowledge-sharing advocate having experience shipping machine learning and software solutions for startups across Asia and Europe.

SKILLS

Programming Languages	Python, C/C++, Java, RISC Assembly
AI/ML Domains	Computer Vision, Natural Language Processing, Graph ML, MLOps
AI/ML Tools	Scikit-Learn, PyTorch, TensorFlow, Keras, AWS SageMaker, 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): 4.9/5.0 – Dean's List
- NUS Overseas Colleges Program, NOC Paris (2023): participated in entrepreneurship program at Université PSL

EXPERIENCE

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

- Developed web domain credibility ranking models with graph neural networks using PyTorch Geometric and NetworkX, built end-to-end ML pipeline, and deployed models in production (making prediction for > 1 million domains) using FastAPI
- Initiated and 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%
- Evaluated and customised social media data scraping tools to provide insights for scientists doing fact-checking task

Research Assistant, SEEDER Group – National University of Singapore **Jan 2022 – Dec 2022**

- Analysed tradeoffs in learning algorithms for spiking fully connected and convolutional neural networks
- Framed architecture and wrote assembly instructions for spiking neural network hardware accelerators on FPGA

Teaching Assistant, School of Computing – National University of Singapore **Aug 2021 – Dec 2022**

- Placed on the Honour List of Student Tutors for Excellence in Teaching with 4.8/5 feedback point (faculty average: 4.2)
- Conducted classes for Master-level IT5003 Data Structures and Algorithms (covered DSA concepts and implementation in Python) and CS2102 Database System (discussed Relational Algebra, ER model, SQL, and Database Normalization)

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 more than 95% accuracy, detecting bad parking behaviours with 0.98 precision, 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)

- Built duplicate question identification models using 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 by utilizing byte-pair encoding, n-grams, MLP, RNN, and transformers accomplishing 76.8 perplexity; augmented generated questions to classification task in a semi-supervised manner

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

- Collected data and created an ML model classifying actions based on IMU data mounted on users' hands attaining 96.14% test accuracy using PyTorch; translated model into high-level synthesis (C++) and ran inference on Avnet Ultra96-V2 board
- Designed architecture, optimized inter-component communication overhead through 20% data packet size reduction

RTOS Project (demo | source code): semi-automated car powered by KL25Z microcontroller, broke course's running time record

- Led team of 4, outlined hardware and software architecture, managed threads using RTX real-time operating system
- Implemented motor movements and communication protocols (Wi-Fi on ESP32 and UART), created web controller interface

AWARDS & HACKATHON

- Champions, Vietnam National AWS DeepRacer League, Dec 2019: formulated reward functions and trained reinforcement learning models loaded to an autonomous car completing a racing lap in 9.625s

EXTRA-CURRICULAR ACTIVITIES

- NUS School of Computing Project Mentor, 2023: guiding 2 Computer Engineering freshmen, offering advice and guidance on academic and non-academic matters, internship, overseas experiences, and university life navigation