BUI DUC THANH

(+65) 8741 0543 ◇ bdthanh@u.nus.edu ◇ linkedin.com/in/buiducthanh2003 ◇ github.com/bdthanh

EDUCATION

National University of Singapore   **Aug 2021 - Expected May 2025**
Bachelor of Engineering, Computer Engineering, Minors in Data Engineering
- Cumulative Average Point: 4.69/5.00 (First Class Honours - Deans' List Sem 1 AY2023-2024)

TECHNICAL SKILLS

Programming Languages	Python, C++ (Proficient), Java (Intermediate)
AI/Machine Learning	Scikit-learn, Pytorch, LightGBM, XGBoost, LangChain, AWS SageMaker, Optuna
Cloud Service	Amazon Web Services
Web Development	React Native, MySQL, SQL
Others	Git, Miro, Docker, Streamlit

PROFESSIONAL CERTIFICATES

- AWS Certified Machine Learning - Specialty
- AWS Certified Cloud Practitioner

EXPERIENCE

AI/ML Engineer Intern, Sabic Asia Pacific Pte Ltd   **Singapore, Jan 2024 - Present**
- Conducted analysis of historical text data from internal ticketing system databases, deriving key insights to improve search ability.
- Developed and implemented an end-to-end ML pipeline for automatic ticket suggestion based on text descriptions, achieving 0.83 F1 score and significantly reducing manual selection errors.
- Built an application that extracts data from Quality Management PDFs to Excel for SAP upload using Optical Character Recognition, eliminating manual processing and saving 10 man-hours/month with 100% accuracy.

Applied AI Engineer Intern, Amaris.AI   **Singapore, Jun 2023 - Aug 2023**
- Assisted in implementing a retrieval augmented generation compliance checking system, by using vector databases and open-source large language models (Llama2-13b, ...).
- Created PDF parser software to extract and organize text from risk assessment documents, yielding clean, high-quality input data for better embeddings and enhanced RAG system performance.
- Engaged in continuous research and testing of cutting-edge open-source models and APIs, fostered effective communication with various teams to ensure seamless system integration.

Teaching Assistant - Software Engineering & OOP, NUS   **Jan 2023 - May 2023**
- Conducted tutorials and facilitated discussions related to Software Engineering and Java language concepts.

PROJECTS

Distracted Driver Detection (team of 5): Built a model to detect distracted behaviours of drivers while driving (Best Project Award for CS3244 Machine Learning in the 22nd NUS School of Computing Term Project Showcase)
- Fine-tuned and conducted experiments on several pre-trained models, including VGG16 and InceptionV3 with Tensorflow on Distracted Driver Dataset, culminating in a noteworthy 0.92 accuracy rate.
- Performed hand segmentation on original dataset to train a sub-model and assisted in ensemble process to increase final model's performance.

English-Vietnamese Neural Machine Translation - Transformer from scratch
- Implemented Transformer architecture with PyTorch from scratch based on "Attention is all you need" paper.
- Trained the model with IWSLT'15 English-Vietnamese dataset containing 133K sentence pair.
- Apply enhancements from Llama 2 architecture, including Rotary Position Embedding, RMSNorm to understand and improve original transformer performance (in progress).

MonKey (pair project - Orbital 2022 - Apollo 11: Advanced Level of Achievement)
- Programmed a cross-platform application for tracking transactions, enhancing financial management with customisable features, developed front-end with React Native and integrated with back-end using Firebase

COMPETITIONS

Tokka Labs Quantitative Challenge 2024 - Kaggle competition Top 6 - $1000 USD consolation prize
- Analyzed historical market data of 10 cryptocurrencies, and developed a machine learning pipeline using LightGBM to forecast 10-minute log returns, achieving an average Pearson correlation coefficient of 0.047.

National AI Student Challenge 2024 - Singapore (Shortlisted to 2nd round - Top 20, final result awaiting)
- Analyzed Petfinder adoption data to identify key factors influencing adoption rates and developed a machine learning pipeline to predict pet adoption speed, enhancing placement efficiency.