BUI DUC THANH

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

EDUCATION

National University of Singapore
Bachelor of Engineering, Computer Engineering, Minors in Data Engineering
- Cumulative Average Point: 4.66/5.00 (First Class Honours)

EXPERIENCE

Software Engineer (Machine Learning), ParallelChain Lab **Singapore, Jun 2024 - Oct 2024**
- Assisted in building backend for KYC SDK, enhancing feature robustness and improving speed by 36% and seamlessly integrating machine learning models into third-party applications.
- Researched on text-to-speech, voice conversion techniques, and speech datasets to enhance our e-KYC application with a voice anti-spoofing feature, and fine-tuned speech models.
- Proposed and implemented an robust audio augmentation pipeline for training process, boosting performance by 24% and securing 2nd place in the ASVspoof5 competition.
- Developed a voice-based gender classification model with 98.26% accuracy.

Machine Learning Engineer, Sabic Asia Pacific Pte Ltd **Singapore, Jan 2024 - Jun 2024**
- Conducted analysis of historical text data from internal ticketing system databases, deriving key insights to improve search ability.
- Implemented an end-to-end pipeline for automatic ticket suggestion based on text descriptions, significantly reducing manual selection errors.
- Built an application that extracts data from Quality Management PDFs to Excel using fine-tuned OCR engine, eliminating manual processing and saving 10 man-hours/month with 100% accuracy.

AI Engineer, 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.
- Created PDF parser software to extract and organize text, yielding clean, high-quality input data for better embeddings and enhanced RAG system performance by 18%.
- Engaged in continuous research and testing of cutting-edge models and APIs, fostered effective communication with various teams to ensure seamless system integration.

PUBLICATION

- *ParallelChain Lab's Anti-Spoofing Systems for ASVspoof 5.* Thien Tran, **Thanh Duc Bui**, Panagiotis Simatis. ASVspoof5 Workshop, Interspeech 2024, August 31, 2024, Kos Island, Greece.

PROJECTS

Laser Tag - AR shooting game (Capstone Project)
- Led the team of 5, collected IMU data on players and trained an motion classification model attaining 99.81% test accuracy; translated model into high-level synthesis (C++) and deployed on Ultra96-V2 board.
- Optimized parallel computing via multithreading, ensuring real-time performance with seamless data acquisition, processing, and transmission.

English-Vietnamese Neural Machine Translation - Transformer from scratch
- Implemented Transformer architecture with PyTorch from scratch based on "Attention is all you need" paper.
- Experimented and trained with IWSLT'15 English-Vietnamese dataset containing 133K sentence pairs.

Distracted Driver Detection: Built a model to detect distracted behaviours of drivers while driving
- 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.

TECHNICAL SKILLS

Programming Languages	Python, C++, Java
AI/Machine Learning	Pytorch, HuggingFace, Optuna, XGBoost, MLflow, NLP, Speech ...
Others	CUDA, Git, Docker, AWS, Gradio, Streamlit, FastAPI, Linux, SQL

COMPETITIONS

Tokka Labs Quantitative Challenge 2024 - Kaggle competition Top 6 - $1000USD prize
- Analyzed cryptocurrencies historical market data, developed alphas, and optimized a machine learning pipeline to forecast 10-minute log returns.