# QUOC ANH BUI

buihanquocanh@gmail.com | linkedin.com/in/alanbui2808 | github.com/alanbui2808

## SKILLS

**Languages:** Python, Ruby, R, Java, JavaScript, C/C++.
**Technologies:** Pytorch, Tensorflow, Rails, Spark, Hadoop, Redis, Postman, AWS, Kafka, Kubeflow, NumPy, Pandas, Grafana, Git, Docker, PostgreSQL, Jenkins, Tableau.

## EDUCATION

Master's in Computer Science at UMass Amherst — Graduation: Jan 2024
**Machine Learning concentrated**. GPA: 3.7
**Relevant Coursework**: NLP - Large Language Models specialized, Machine Learning, Neural Nets, Advanced Algorithms, Database, Optimization, Networking and Security, Information Retrieval, Hypothesis Testing and Business Analytics.
**Awards**: UMass CICS **Bay State Program** Scholarship and **Dean's List**.

## EXPERIENCE

**Research Assistant/Participant @ CICStep Program** | *Python, Pytorch* — **UMass CICS**, Feb - Jun24

- Built data processing pipeline using **Elasticsearch** for efficient indexing and embedding of BioLab data. Developed a robust **RAG** framework with **DPR** integration using PyTorch and Elasticsearch, resulting in **30**% better relevance and **15**% higher user engagement.
- Intergrated multiple LLM models (**Llama-2**, **Mistral** and **StableLM**) on RAG outputs, optimizing language understanding and boosting response accuracy and satisfaction metrics.

**Data Analytics Intern** | *Tensorflow, Pandas* — **Viet Pho Television Network LLC**, Jul 22 - Feb 2023

- Collaborated with the IT department to gather viewer behaviors and preferences. Utilized Graph API to gather interactions from Facebook posts. Built a data processing pipeline for over **5K** audiences and 100 Facebook posts using **Pandas** and **NumPy**.
- In a team of 4, employed XGBoost model using **Tensorflow** to predict audience preferences, resulting in a **15**% increase in viewer engagement.

**Backend & ML Engineer Intern** | *Ruby on Rails, Pytorch* — **Ascenda Loyalty**, May - Nov 2020

- Worked in the E2E Rewards Program to develop routing and retrieval feature for the main orchestrational service using **RoR**, **Kafka**, **RESTful APIs**, **GraphQL**, and **JWT**, achieving an increase of **15**% response time. Supporting **300k** daily active users from **4** international banks.
- Integrated **Redis** and **AWS ElastiCache**, reducing data access latency by **20**% for faster data retrieval and scalability. Integrated real-time monitoring using **AWS CloudWatch** and **Grafana**, ensuring system health and stability.
- In team of 4, developed **Purchase Eraser** service that enables users to erase global transactions using accrued point rewards. Optimized with **GraphQL**, **Redis** caching and **PostgreSQL** for efficient communication, retrieving user-specific data from internal services for swift request handling.
- In team of 4, developed **APIs** for the **Giftcards Handler** service to establish an unified data source for other services using **Hanami**. Integrated Hadoop for processing 50,000 giftcards from 1500+ suppliers, optimized storage using PostgreSQL.
- In team of 10, **fine-tuned** a **RoBERTa-Large 355M** fraud detection model over **500k** collective activities from internal banks and Ascendas, using **Pytorch**, **PySpark** and **Pandas**. Enhanced model capability with aggregated sequential transactional data and encoded users behaviors, resulting in **70 ROC-AUC**.
- Developed a service encapsulating the model into the existing ecosystem, enhancing fraud detection during client/user request handling using RoR. Streamlined deployment with **Kubeflow** for continuous monitoring and scalability.
- Conducted **Canary Deployment** that directed **15**% of the traffic to monitor the model's performance and stability via **AWS SageMaker**.
- Migrated databases and configured multi-tenant endpoints for up to 5 tenants, maintaining APIs simplicity and performance.

**Research Assistant** | *Python, Pytorch* — **UMass CICS**, Jun - Aug 2019

- Trained a Bayesian Knowledge Tracing hidden markov model for Professor Andrew Lan's tutoring system, utilizing Baum-Welch algorithm to re-estimate students' learning parameters.
- Successfully scaled the model to incorporate input from a batch of students' answers sequences, achieving a 65% increase in accuracy.

## PROJECTS

**Recipe Infusion** | *Python, Pytorch*

- Led a team of 4 in **fine-tuning DistilGPT2** on the aggregated RecipeNLG and RecipeBox of 2M recipes to generate recipes.
- Implemented **Style Transfer** by fine-tuning**T5** utilizing **backtranslation** to infuse different personas' styles from respective corpora of Taylor Swift, Donald Trump and Shakespeare.
- Achieved 0.34 **BLEU** on recipe generation and 65% successful rate (**human evaluation**) on style infusion.

**RSNA-MICCAI Brain Tumor Radiogenomic Classification**| *Python, Tensorflow*

- Led a team of 3, trained a **3D CNN** model on dataset of 100 patients' structural **multi-parametric** of 20K MRI scans to detect MGMT genetic sequence that signals potential of developing brain tumor for **BioNLP** UMass.
- Achieved relatively good accuracy score of **0.56** through rigourous optimization and data processing.

**Coding Interview Problems classification** | *Python, Pytorch*

- Fine-tuned **BERT-base** on self-collective and annotated dataset of 300 coding problems from Leetcode, Hackerrank, etc to achieve **75**% **ROC-AUC** in problems' difficulty classification that truly reflects interviewee's experience.