Nga (Jane) Vu

623-275-1633 | vuthiquynhnga4@gmail.com | linkedin.com/in/nga-vu | github.com/ngavu2004

Education

Arizona State University **Expected Dec 2026**
Master in Computer Science (GPA: 3.90) *Tempe, AZ*
Courses: Distributed Software Development, Cloud Computing, Operating Systems, Data Structures and Algorithms

Technical Skills

Cloud: AWS Cloud (EC2, S3, Lambda, ECR, SQS, Sagemaker, Bedrock)
Languages: Java, Python, C/C++, SAS, SQL (Postgres), JavaScript, HTML/CSS, R
Frameworks/ Developer Tools: React, FastAPI, Git, Docker
Database: SQL/SQLite, PostgreSQL, MongoDB, Firebase

Professional Experience

Software Engineering Intern Feb 2024 - Present
Dr.George Runger's Laboratory - Arizona State University *Tempe, AZ*
- Architected a serverless **RAG chatbot** system with **AWS API Gateway, Lambda, and Neo4j** database for TIPQIC, reducing backend latency by **60%** and scaling to serve 100+ monthly user queries securely..
- Created an **ETL pipeline** on 20GB of text data utilizing **Langchain**, and **Python** on **Llama3** to automate assessments, resulting in reduced operational costs by approximately 30% in a 200-clinic evaluation program.
- **Hosted weekly sessions** about AI automation for clinical data to a team of **10 non-technical stakeholders** from the AZ Health Department, enabling informed decision-making and successful adoption.

Web Developer Feb 2023 - Feb 2024
International Logistics and Productivity Improvement Laboratory - Arizona State University *Tempe, AZ*
- Engineered comprehensive web solutions leveraging **React** with **MongoDB** integration on **AWS Amplify** to deliver **data visualization** leading to heightened collaboration in local farmers community by 30% to 50 users.
- Deployed a machine-learning model to predict market crop yield using **Python**, **TensorFlow**, **SQL** VBA and **ETL** process, achieved 90% accuracy.

Software Engineering Intern May 2023 – Aug 2023
GeoComply *Ho Chi Minh City, Vietnam*
- Migrated the frontend codebase from **Laravel** to over **30 modular ReactJS components** to improve compatibility with mobile apps and new API endpoints by 60%.
- Engineered an **MVC** framework using **PHP** to safely access customer information stored in **Microsoft SQL Server**, elevating the front-end display capabilities and testing efficiency by 50%.
- Optimized complex SQL queries and introduced indexing strategies, reducing data retrieval times by 30% and improving overall system responsiveness.

Projects

Automated Job Tracker — automatejobtracker.com | *Django, PostgreSQL, GCP, AWS, BERT Classifier*
- Led a team of 3 people to build an end-to-end job application tracker that automates Gmail email parsing and updates a connected Google Sheet, reducing manual tracking time from **4 seconds/email to 0.004/email**.
- Fine-tuned a BERT-based transformer model on **500+ labeled emails**, improving classification accuracy from **89% to 94%**, and significantly reducing false positives from job ads.
- Implemented asynchronous job queue using **Celery** and **RabbitMQ** to handle long-running tasks like Google Sheet updates, improving app responsiveness and reliability for concurrent users by 60%.

KeenKaraoke: Youtube URL to Karaoke (SunHacks Hackathon) | *AWS, Python, React, MongoDB, Redis*
- Designed a website to transform YouTube URL to karaoke experience with **ReactJS**, **Django**, and **AWS EC2**, **S3**.
- Integrated AWS **ElastiCache** for **Redis**, leading to a 30% reduction in audio retrieval costs from AWS **S3** and boosting the **FastAPI** application's response time to under 200 milliseconds.
- Implemented **OpenAI Whispers API**, **AWS Bedrock**, and **Google speech-to-text AI** to handle speech-to-text tasks with an accuracy of 95%.

Master Vault (Won HackPrinceton Best self-hosted model prize) | *OpenAI Whisper, Gemini, SQLite*
- Led a team of 3 members to develope an AI-powered podcast system that converts short-form videos (TikTok, YouTube Shorts) into structured, long-form learning content, processing **50+ videos per minute**.
- Optimized **Google Gemini, Microsoft Phi-3, and OpenAI Whisper** performance with dynamic model routing system, achieving **95% transcription accuracy** and **30%** faster podcast generation.
- Improved backend server efficiency using **FastAPI with SQLite caching**, cutting response times by 50% compared to traditional databases.