Hieu Minh Nguyen

EDUCATION

Tampa, Fl	University of South Florida	2021 - Expected May 2026

- Bachelor's Degree / Master's Degree in Computer Science. GPA: 3.8/4.

SKILLS

- **Programming Language**: Python, C, C++, C#, Java, JavaScript, HTML5, CSS, SQL, Bash.
- **Technologies and Tools**: MySQL, PostgreSQL, NoSQL, Git/GitHub, GitHub Action, AWS, Google Cloud, Azure, Docker.
- **Web development Technologies**: Node.js, Vite.js, Express.js, React.js, FastAPI, Bootstrap, React Native, Flask, Next.js.

PROFESSIONAL EXPERIENCE

Research Assistant	University of South Florida	May 2024 – Present

Publication: "Global is Good, Local is Bad?": Understanding brand Bias in LLMs.
- Conducted comprehensive research on bias across many categories, such as age or gender, within Large Language Model (LLM).
- Built a specialized library in **Python** for systematic evaluation of **100,000** datasets in LLMs, significantly reducing manual effort.
- Collaborated with a team of researchers on studies focusing on mitigating bias in human health records using machine learning.
- Researched techniques to optimize LLM code generation for **complex data structures, algorithms**, competitive programming.

AI/ML Software Engineer (Intern)	Resilience, Inc.	Sep – Dec 2023

Technologies: AWS, React Native, Flask, Synk, Figma, Linux, Python.
- Engineered a mindfulness app, resulting in enhanced emotional self-awareness across diverse user demographics.
- Executed design prototypes using Figma, further developing **5+** frontend features via React Native.
- Implemented a machine learning API endpoint with Flask (Python) on AWS EC2, leading to an **80%** reduction in response latency.
- Coordinated with cross-functional teams, include engineering and business for smooth production releases.

AI/ML Developer (Intern)	Resilience, Inc.	June – Aug 2023

Technologies: Keras, Numpy, Pandas, sklearn, matplotlib, TensorFlow, Python.
- Preprocessed and visualized **7000+** audio samples to extract vital features and improve the quality of training data.
- Built a CNN model using Keras **(**TensorFlow**)** from scratch, achieving a **72%** emotion classification accuracy.
- Streamlined machine learning processes by creating a pipeline for automated data handling and model training.
- Finetuned Whisper by **multiprocessing** through **parallelization** for CPUs to speed up translation time by **5.9x**.

PROJECTS

Talk To Listen (talktolisten.com) | FastAPI, Python, React Native, JavaScript, REST APIs, SQL, Firebase, Redux, CI/CD, AWS, Azure.
- Built a **SaaS** app with **1000+** active users and secured **$3000** in sponsorship from Microsoft, OpenAI, and AWS Startup Programs.
- Managed all stages from front-end, database, API design, and back-end, for a **codebase** exceeding **3000** lines.
- Created **scalable distributed** system architecture using Azure Application Gateway, Azure Load Balancer, Azure Virtual Machines.
- Constructed a RESTful APIs server with FastAPI (Python) capable of handling 30+ APIs, and 3 machine learning endpoints, deployed on AWS EC2 using Docker, and integrated with GitHub Actions for automated CI/CD.
- Created and published an **open-source** dataset of **3M** tokens (nearly **2000** downloads monthly) and fine-tuned an LLM model using supervised learning, quantization method (QLoRA), and Pytorch, achieving an accuracy of **85%**.

JobsDreamer (jobsdreamer.com) | Scrapy, Selenium, Python, Google Gemini 1M API, GPT4o, Llama3, Proxy, Azure, AWS.
- Launched a platform that assists **hundreds** of students in finding internships, automatically sending them emails with newly all posted relevant internship opportunities on the internet within the last 24 hours.
- Established a comprehensive job data processing pipeline utilizing **latest** AI tools, AWS EventBridge, AWS Step Functions, AWS Fargate, and AWS Lambda to **automate** daily web scraping, data preprocessing, and email notifications.
- Deployed containerized **microservices** for web scraping across multiple job sites, leveraging Python, Scrapy, and Selenium.
- Automated a scalable and efficient Python pipeline with **advanced LLMs** to handle **1000+** of new **raw data** every day.

BullBot (bullbot.tech) | Flask, AWS, Vite.js, Express.js, Node.js, Pytorch, Langchain, Python.
- Pioneered a full stack chatbot web app answering queries across **1500+** Uni of South Florida websites with natural language recognition, providing precise sources for user clarity, especially for parents less adept at online searches.
- Initiated a Retrieval Augmented Generation (RAG) system, integrating Vector Store for efficient data management.
- Employed Depth First Search **Algorithm** with Beautiful Soup for comprehensive web link scraping.
- Orchestrated a **microservice** architecture leveraging Vite.js, Express.js/Node.js, and various AWS services.

LEADERSHIP

Software Lead	Association for Computing Machinery – USF Chapter	November 2023 – Present

- Leading a team project of 8 by creating detailed project plans, functional requirements, and documentation.
- Hosting technical workshops and mentoring other members in preparing and delivering their workshops, fostering a collaborative learning environment.