



# Hieu Minh Nguyen

## EDUCATION

Tampa, FL	University of South Florida	2021 - Expected May 2026
<ul style="list-style-type: none"><li>Bachelor's / Master's Pathways Degree in Computer Science. GPA: 3.8/4.</li></ul>		

## SKILLS

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

## PROFESSIONAL EXPERIENCE

Research Assistant	University of South Florida	May 2024 – Present
Publication: “Global is Good, Local is Bad?”: Understanding brand Bias in LLMs.		
<ul style="list-style-type: none"><li>Conducted comprehensive research on bias across many categories, such as age or gender, within Large Language Model (LLM).</li><li>Built a specialized library in <b>Python</b> for systematic evaluation of <b>100,000</b> datasets in LLMs, significantly reducing manual effort.</li><li>Collaborated with a team of researchers on studies focusing on mitigating bias in human health records using machine learning.</li><li>Researched techniques to optimize LLM code generation for <b>complex data structures, algorithms</b>, competitive programming.</li></ul>		
AI/ML Software Engineer (Intern)	Resilience, Inc.	Sep – Dec 2023
Technologies: AWS, React Native, Flask, Synk, Figma, Linux, Python.		
<ul style="list-style-type: none"><li>Engineered a mindfulness app, resulting in enhanced emotional self-awareness across diverse user demographics.</li><li>Executed design prototypes using Figma, further developing <b>5+</b> frontend features via React Native.</li><li>Implemented a machine learning API endpoint with Flask (Python) on AWS EC2, leading to an <b>80%</b> reduction in response latency.</li><li>Coordinated with cross-functional teams, include engineering and business for smooth production releases.</li></ul>		

AI/ML Developer (Intern)	Resilience, Inc.	June – Aug 2023
Technologies: Keras, Numpy, Pandas, sklearn, matplotlib, TensorFlow, Python.		
<ul style="list-style-type: none"><li>Preprocessed and visualized <b>7000+</b> audio samples to extract vital features and improve the quality of training data.</li><li>Built a CNN model using Keras (TensorFlow) from scratch, achieving a <b>72%</b> emotion classification accuracy.</li><li>Streamlined machine learning processes by creating a pipeline for automated data handling and model training.</li><li>Finetuned Whisper by <b>multiprocessing</b> through <b>parallelization</b> for CPUs to speed up translation time by <b>5.9x</b>.</li></ul>		

## PROJECTS

<b>Talk To Listen</b> ( <a href="http://talktolisten.com">talktolisten.com</a> )   FastAPI, Python, React Native, JavaScript, REST APIs, SQL, Firebase, Redux, CI/CD, AWS, Azure.		
<ul style="list-style-type: none"><li>Built a mobile AI chat app for seamless interaction with unique AI characters, each featuring distinct voices, personalities, stories.</li><li>Grew the app to <b>1000+</b> active users and secured <b>\$10,000+</b> in sponsorship from Microsoft, OpenAI, and AWS Startup Programs.</li><li>Managed all stages from front-end, database, API design, and back-end, for a <b>codebase</b> exceeding <b>3000</b> lines.</li><li>Created <b>scalable distributed</b> system architecture using Azure Application Gateway, Azure Load Balancer, Azure Virtual Machines.</li><li>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.</li><li>Created and published an <b>open-source</b> dataset of <b>3M</b> tokens (nearly <b>2000</b> downloads monthly) and fine-tuned an LLM model using supervised learning, quantization method (QLoRA), and Pytorch, achieving an accuracy of <b>85%</b>.</li></ul>		
<b>JobsDreamer</b> ( <a href="http://jobsdreamer.com">jobsdreamer.com</a> )   Scrapy, Selenium, Python, Google Gemini 1M API, GPT4o, Llama3, Proxy, Azure, AWS.		
<ul style="list-style-type: none"><li>Launched a platform that assists <b>hundreds</b> of students in finding internships, automatically sending them emails with newly all posted relevant internship opportunities on the internet within the last 24 hours.</li><li>Established a comprehensive job data processing pipeline utilizing <b>latest</b> AI tools, AWS EventBridge, AWS Step Functions, AWS Fargate, and AWS Lambda to <b>automate</b> daily web scraping, data preprocessing, and email notifications.</li><li>Deployed containerized <b>microservices</b> for web scraping across multiple job sites, leveraging Python, Scrapy, and Selenium.</li><li>Automated a scalable and efficient Python pipeline with <b>advanced LLMs</b> to handle <b>1000+</b> of new <b>raw data</b> every day.</li></ul>		
<b>BullBot</b> ( <a href="http://bullbot.space">bullbot.space</a> )   Flask, AWS, Vite.js, Express.js, Node.js, Pytorch, Langchain, Python.		
<ul style="list-style-type: none"><li>Pioneered a full stack chatbot web app answering queries across <b>1500+</b> Uni of South Florida websites with natural language recognition, providing precise sources for user clarity, especially for parents less adept at online searches.</li><li>Initiated a Retrieval Augmented Generation (RAG) system, integrating Vector Store for efficient data management.</li><li>Orchestrated a <b>microservice</b> architecture leveraging Vite.js, Express.js/Node.js, and various AWS services.</li></ul>		

## LEADERSHIP

Software Lead	Association for Computing Machinery – USF Chapter	November 2023 – Present
<ul style="list-style-type: none"><li>Leading a team project of 8 by creating detailed project plans, functional requirements, and documentation.</li><li>Hosting technical workshops and mentoring other members in preparing and delivering their workshops, fostering a collaborative learning environment.</li></ul>		