

# Minh (Tommy) Nguyen

484-620-7355 | [tnguye59@villanova.edu](mailto:tnguye59@villanova.edu) | [linkedin/in/tommynguyen04/](https://linkedin/in/tommynguyen04/) | [github/Tommy0201](https://github/Tommy0201) | [tommy0201.netlify.app](https://tommy0201.netlify.app)

## EDUCATION

### Villanova University

Expected Dec 2025

GPA: 3.97/4.00

Bachelor of Science in Computer Science - Honors Degree, Business Minor

- **Awards:** SHPE 2025 National Conference Scholarship (100%), Tapia 2024 Conference Scholarship (100%), Connelly-Delouvrier Full Tuition Scholarship, AI-JAM US Silicon Valley (Gold medal)
- **Coursework:** Machine Learning, Large Language Models, Data Science, Statistics Methods, Database Systems, Linear Algebra, Discrete Mathematics, Operating Systems, Data Structures and Algorithms, Software Engineering
- **Research Paper:** Improving Large Language Models Linguistic Diversity published by **NAACL 2025**.

## EXPERIENCE

### Olli Technology

May 2024 – Aug 2024

Remote

AI Engineer Intern – Python, LangGraph, LangChain, Flask, TypeScript, React, gRPC

- Co-architected a personalized AI voicebot with a 2-person team using multi-agent **LangGraph workflows**
- Engineered an AI-powered image recommendation system for a marketing assistant platform by integrating **Unsplash API** and **DALL-E API**, boosting **user engagement** by **30%**, and reducing **design time** by **40%**
- Delivered a **production-ready** prototype that catalyzed a **nationwide pilot program** and secured a partnership with Malaysia's Minister of Education, projected to impact **5M+ students**

### Product Manager Accelerator

Sep 2024 – Dec 2024

Boston, MA

Data Scientist Intern – Python, Pandas, NumPy, Scikit-learn

- Developed a recommendation engine leveraging **vector similarity**, **one-hot encoding**, and **rank-based algorithms** to amplify visibility of BIPOC-owned small businesses, boosting **1000+ impressions** monthly
- Engineered a vectorized database by extracting and processing data from **1,000** business profiles into **20** feature embeddings with **GooglePlaces API**, reducing **query latency** by **80%** for downstream analysis
- Optimized SEO for **100 businesses**, boosting **average search ranking** by **40%** compared to baseline

### Villanova University

May 2023 – May 2025

NLP Research Assistant – Transformer, NLP, Pandas, BeautifulSoup, Concurrent, HuggingFace

Villanova, PA

- Fine-tuned 3 large language models (DeltaLM, ByT5, NLLB) for translation tasks using benchmark datasets of **700,000 sentences**, outperforming **GPT-4o** with **BLEU-score** improved by **26%, 18% and 9%** respectively
- Published 2 quality datasets of **150,000 parallel sentences** in poor-resourced languages, Nigerian Pidgin and Igbo, annotated within **9 months** by **15 native speakers** and undergone thorough quality validation
- Web-scraped and preprocessed **2,000,000 sentences** from BBC and TedTalks using **concurrent programming**

### SoFi Technologies

May 2025 – Aug 2025

Software Engineer Intern – Spring Boot, ReactJS, PostgreSQL, GraphQL, SourceGraph

San Francisco, CA

- Developed an AI suggestion tool to parse tickets, retrieve historical fixes, and recommend SQL commands, reducing **average resolution time** by **35%** and significantly boosting engineers efficiency
- Engineered a full-stack data change tool with bulk insertion and dynamic pre-validation, reducing **turnaround time** by **85%**, saving **150+ engineering hours** annually, and preventing error cascades from faulty data entry

## PROJECTS

### Facial Emotion Recognition Model | Python, TensorFlow, Keras, Google Cloud | [link](#)

Aug 2024 – Dec 2024

- Developed **5 models** to classify 7 basic emotions using RAFDB and AffectNet datasets of **43,000 images**
- Implemented a **custom CNN model** with fine-tuned layers and attention mechanisms, achieving **70% accuracy**

### DocScribe AI | OpenCV, EasyOCR, LangChain, Python, Flask, Docker, AWS | [link](#)

May 2024 – Present

- Built a **RAG** chatbot with **ChromaDB** for vector-based file retrieval and multi-session chat interactions
- Provided **image-to-text** extraction using Contour Detection, Warped Transformation, and **EasyOCR** model

## TECHNICAL SKILLS

**Languages:** Python, Java, SQL, PostgreSQL, R, JavaScript, TypeScript, C/C++, HTML/CSS

**Frameworks:** ReactJS, VueJS, Flask, Spring Boot, REST API, gRPC

**Developer Tools:** Git, Gitlab, Docker, AWS, DigitalOcean, Firebase, MongoDB, MySQL, VS Code, RStudio, IntelliJ

**Libraries:** PyTorch, TensorFlow, Transformers, Scikit-learn, LangGraph, LangChain, OpenCV, Pandas, NumPy