

Quang Hoang

Fairfax, VA | qhoang@wm.edu | (571) 394-0445 | [in/quanghoang1](https://in.quanghoang1) | github.com/theantigone | cs.wm.edu/~qhoang
U.S. Citizen | Active Public Trust | Eagle Scout | Community Service Awards

Summary

AI Software Engineer specializing in **Python**, **AWS infrastructure**, **RAG pipelines**, and **agentic workflows**

Education

William & Mary

Bachelor of Science in Computer Science | Major GPA: 3.49

Williamsburg, VA

Aug 2022 – Jan 2026

Experience

AI Engineer

Lumiture, LLC

Apr 2026 – Present

Washington, DC

- Accelerated production deployments by building 153-case **LiveKit** regression voice testing pipeline with **DeepEval**
- Eliminated session hangs under peak load by debugging async cleanup deadlock in LiveKit worker loop
- Achieved 100% block rate of out-of-scope inquiries through **LangGraph** system prompting and **pytest** regression suites
- Routed callers to correct live agent queues by porting legacy call flow logic into multiframe LangGraph orchestration deployed via **Terraform** and **Kubernetes** on **Amazon EKS**

AI Trainer

Handshake AI Fellowship

Dec 2025 – Apr 2026

Remote

- Improved conversational LLM response quality across ~1000 multi-modal evaluations by analyzing output accuracy, clarity, and depth
- Supported prompt development and evaluation tasks by conducting independent research

Agile Software Engineer (Academic Capstone)

William & Mary

Aug 2025 – Dec 2025

Williamsburg, VA

- Delivered 4 sprint milestones for 6-person Agile team by tracking and managing 54 feature tickets in **Jira**
- Maintained codebase quality in **GitHub** across 53 reviewed PRs by adhering to feature-branch workflows
- Automated Qualtrics dataset reconciliation alongside manual data formatting by building a containerized **Django** and **Docker** survey analysis platform
- Secured regression safety for critical data parsing views by writing 33 Django unit tests

Software Developer Intern

Soft Tech Consulting

May 2025 – Aug 2025

Fairfax, VA

- Automated internal company knowledge retrieval by deploying LLM chatbot with OneDrive RAG in Copilot Studio
- Streamlined internal contract management by resolving CLIN and task order modifications dependencies by designing relational **Dataverse** schema with **Power Fx** in Teams

Projects

Static Site Generator | *Python, HTML/CSS, CI/CD*

May 2025 – Jun 2025

- Automated web deployment into 1-step Markdown-to-HTML build pipeline by building static site generator in **Python**
- Ensured HTML generator stability and parser reliability by writing 100 unit tests via **unittest**

Generative AI & NLP Engineering Portfolio | *PyTorch, Hugging Face, Pandas, OpenAI*

Jan 2025 – May 2025

- Automated Python code repair and if-statement completion, achieving 0.88 **CodeBLEU** score, by fine-tuning CodeT5-Small model with **Hugging Face Trainer**
- Evaluated LLM code translation and bug fixing across 22 tasks via **BLEU** and CodeBLEU metrics by engineering and benchmarking 5 prompting strategies on GPT-4 and Codestral models
- Engineered Java code completion system with low baseline perplexity of ~177 by developing **N-gram Probabilistic Language Model** in **Python**

Technical Skills

Languages: Python, JavaScript/TypeScript, SQL, C, Java, HTML/CSS, Power Fx

Frameworks: LangChain/LangGraph, DeepEval, LiveKit, Django, Flask, PyTorch, Hugging Face Transformers, pytest, unittest

Developer Tools: Git/GitHub/GitLab, AWS Bedrock, Terraform, FAISS, Docker, Kubernetes, LiveKit, Jira, uv

Libraries: Deepgram, Boto3, Azure OpenAI