JEREMY DO

740-915-3932 | do_g1@denison.edu | /in/jeremy-do/ | /github.com/dohoanggiahuy317 | portfolio

EDUCATION

Denison University   Granville, OH   Graduation: May 2026
- **Bachelor of Science** in Computer Science and Applied Mathematics (**Major GPA: 4.0**)
- **Coursework**: Cloud services (AWS), Data Systems, Microservices and Serverless (IBM), NLP with Deep Learning (Stanford), Computer Systems
- **Leadership:** Microsoft Student Ambassador, Denison Computer Science Department Academic Fellow, Teaching Assistant

AWARDS

National Mathematics Olympiad 2019: Third Prize in the Vietnam National Mathematics Olympiad (USAMO equivalent)
Ohio Competitive Programming Contest 2023 & 2024: First Prize among 45 programming teams from 4 colleges in Ohio
ICPC Regional ECNA contest 24: Top 38 among 100+ programming teams from East Central North America

SKILLS

Development Tools: C/C++, Java, JavaScript, Python, React, Assembly, Spring boot, Firebase, Postman, CRUD, gRPC, Jmix, Kafka, Redis
Tools: Docker, Kubernetes, MySQL, PostgreSQL, Monasca, AWS S3/Lambda/EC2, DynamoDB, Microsoft Azure Functions, .NET , RabbitMQ

EXPERIENCE

Capital One   May 2025 – Present
Software Engineer Intern   Richmond, VA
- Bank Tech team

Deloitte   May 2024 – August 2024
Software Engineer Intern   Hanoi, Vietnam
- Enhanced the security of **4,000+** daily bank transactions by leveraging Jmix and Spring Boot to develop monitoring and protection features
- Optimized SQL queries, reducing loading time of daily transaction reports by **2 min (80%)** by creating index and using aggregation functions
- Implemented and deployed an Algolia search service with multiple replicas, improving reliability and peak throughput of the search service
- Implemented load balancing to distribute incoming requests across multiple Docker containers and separated different services into dedicated containers, increasing throughput by **35% (~1.5k** requests per second)

Above Data, Inc   January 2024 – April 2024
Software Engineer Intern   Granville, OH
- Designed a RAG application with LLMs to enhance UX by generating natural responses to customer queries about info on **500+** retail brands
- Automated data scraping using LangChain RecursiveURL and lazy loading, reducing retrieval time from **3,000+** websites by **7 hours (40%)**
- Developed RESTful API to retrieve relevant documents from the Qdrant vector database and transmit them to the Ollama server via gRPC

FPT Software   July 2023 – September 2023
Software Engineer Intern   Hanoi, Vietnam
- Designed a soft deletion feature to facilitate data center procedure correction using C++, Amazon RDS, and DynamoDB
- Decreased average response time from **10 sec** to **7 sec** per client request by utilizing RabbitMQ for async processing and Redis for caching

VinBrain x Microsoft Smart Health Lab   May 2022 – September 2022
Software Engineer Intern   Hanoi, Vietnam
- Optimized scheduling system for **100** patients and doctors during the COVID pandemic by developing a Spring boot microservice app
- Enhanced security at the API Gateway by utilizing HS256 JSON Web Token encryption and sync communication with distributed tracing

PROJECTS

A Novel Pipeline for Creating Machine Learning Dataset | Paper: *https://arxiv.org/abs/2309.05103v1*
- Reduced mental health risks for **1,000,000+** low-paid data filtering workers in India and developing countries by designing a pipeline for creating datasets for Gen AI with a research group of **5**
- Proposed a multi-threaded pipeline to generate a high-quality dataset of **77,000** entries for training LLMs on low-resource languages
- Reduced the expense of crowd-sourced annotator datasets by **50% ($20k)** through randomized matching of query-context

Snowflake RAG 'n' ROLL 2025 – LLM Diversity Holiday Insights Assistant | *GitHub*
- Fostered diversity and inclusivity among **1,000+** international students at my university by leading a team of **5** to develop a microservice RAG app on AWS S3 that delivers information about signature holidays across various countries
- Designed a LangChain contextual compression layer, reducing token usage by **2,000**/query and LLM response time by **2 min**/request

Hackonomics Hackathon 2025 - Student Teamwork Management System | *GitHub*
- Created a light SQL engine using C++ and distributed Parallel Buffer Pool for managing **3,000** teamwork assignments at my college
- Designed a scalable Query Execution Engine capable of parallel and distributed computing to handle the high request load in exam seasons

TikTok TechJam Hackathon 2024 - Image 2 LaTex converter | *GitHub*
- Empowered **500** underrepresented students in rural Vietnam with accessible math education by collaborating with a team of **3** to develop a full-stack app that seamlessly converts images to LaTeX, leveraging AWS S3 for scalability and efficiency