

# VARUN ANUNSHEEL VIKRAM SHA

vikramsha.v@northeastern.edu | +1 (857) 218 009 | <https://github.com/varun7778>  
[www.linkedin.com/in/varun-anusheel7778](https://www.linkedin.com/in/varun-anusheel7778) | <https://varun7778.github.io/>

## EDUCATION

**Northeastern University, Khoury College of Computer Sciences, Boston** May 2025  
Master of Science in Artificial Intelligence CGPA: 3.9/4.0  
Coursework: Pattern Recognition and Computer Vision, Foundations of AI, Machine Learning, Programming Design Paradigm  
**Manipal Institute Of Technology, Manipal, India** Oct 2021  
Bachelor of Technology in Computer and Communication Engineering with a minor in Soft Computing CGPA: 7.98/10  
Coursework: Computer Vision, Artificial Intelligence, Neural Network and Fuzzy Logic, Natural Computing, Cloud Computing, Data Mining & Predictive Analysis, Reinforcement Learning, Parallel Programming, Data Structures, Algorithms, Database Systems

## SKILLS

**Languages:** Java, Python, C++, C, Dart, R, SQL  
**Libraries & Frameworks:** PyTorch, Langgraph, Tensorflow, Keras, Cuda, Numpy, Spicy, Scikit-Learn, OpenCV, NodeJs, Spacy, NLTK, Oracle DB, Langchain, Llamaindex, Hugging Face, Ollama, PostgreSQL  
**Cloud Technologies:** Google Cloud Platform, AWS EC2, AWS Sage Maker, Heroku, AWS Lambda, Azure  
**Skills:** AI, CV, NLP, Machine Learning, Deep Learning, Data Mining, Generative AI, LLMs, Docker

## PROFESSIONAL EXPERIENCE

**XNode, Delaware, USA** May 2024- Present  
*AI Intern Engineer*

- Developed 3 ReAct LLM agentic bots using LangGraph and PGVector to boost query accuracy by 35% and automate processes, increasing response efficiency by 40%, speeding issue resolution, and improving time for creating and resolving tickets.
- Automated BRD generation with an info-gathering bot using RAG and LLM, cutting manual effort by 50%. This ensured more consistent, accurate, and faster delivery of BRD specs, streamlining project initiation and management.

**General Electric, Bengaluru, India** Jun 2021- Aug 2022  
*DevOps Engineer - Sourcing & Logistics Team*

- Spearheaded a team of 5 to build a chatbot using AWS, Oracle BI publisher and Postman and integrated it into Microsoft Teams to directly retrieve user required data from the Oracle database resulting in a ~8% reduction in query tickets.
- Investigated and resolved integration bugs in OTM by analyzing Tableau reports and automating issues, resulting in a 10% productivity boost and decreasing tickets raised by ~18%.

## PROJECTS

### LLM Medical Langchain Project

- Developed the project using LangChain to optimize patient-doctor interactions through Monte Carlo Tree Search (MCTS), resulting in enhanced decision-making and improved patient outcomes.
- Integrated a FAISS-based retrieval system within the project, utilizing HuggingFace models to generate embeddings of medical text data for efficient retrieval of relevant patient cases and medical literature.

### RAG LLM Chatbot

- Developed a chatbot using Large Language Models (LLMs) and Retrieval-Augmented Generation (RAG) technology. This resulted in more accurate and contextually relevant answers to user inquiries along with the source of the answers.
- Implemented a knowledge base integration within the LLM framework, enabling the chatbot to access and retrieve information 15% faster compared to traditional retrieval methods.

### MLOps with CI/CD Pipeline

- Developed a deep learning-based object detection and segmentation pipeline from scratch with an overall accuracy of 87% that can identify, track, and count multiple objects in real time using Mask R-CNN.
- Deployed on Google Kubernetes Engine (GKE) by creating a Streamlit app and then Dockerizing it. Automated the build and deployment process by linking Cloud Build to the GitHub repository, where GitHub is used for version control. The deployed application can be accessed through an endpoint for testing and use.

### MLflow with AWS Deployment

- Developed and deployed a kidney disease classification project using ML, deep learning frameworks like TensorFlow, and workflow management tools such as MLflow and DVC, demonstrating proficiency in model development and version control.
- Implemented AWS CI/CD deployment using GitHub Actions, configuring EC2 instances, Docker containers, and ECR repositories, showcasing expertise in cloud computing, DevOps practices, and automated deployment pipelines.

## ACHIEVEMENTS & ACTIVITIES

- Served as a Teaching Assistant for the Network Security course during the Summer 2024 semester.
- Secured 2<sup>nd</sup> position in the annual guild competition held by GE for developing an AR application that displays models of all the parts and working of various production turbines and production centers in real time on the user's screen.
- Headed a student project called Vision that works to build a Mixed Reality (MR) Headset by getting sponsorships and actively participated in recruiting new team members.