

# JITEN BHALAVAT

✉ jbha0504@umd.edu | ☎ (240) 481-5543 | LinkedIn | GitHub | jitенbhalavat.com

## EDUCATION

### University of Maryland, College Park

Master of Science in Applied Machine Learning

Aug 2024 - Expected May 2026

College Park, Maryland

### Charotar University of Science and Technology (CHARUSAT)

Bachelor of Technology in Information Technology

Sept 2020 - May 2024

Gujarat, India

## SKILLS

**Languages and Databases:** Python, C/C++, JavaScript, SQL, MySQL, MongoDB, Pinecone, Qdrant, PostgreSQL

**Frameworks and Libraries:** PyTorch, TensorFlow, Keras, HuggingFace, Scikit-learn, NumPy, Pandas, Transformers

**Generative AI:** Large Language Models (LLMs), AI Agents, RAG, Prompt Engineering, Fine-tuning

**Tools & DevOps:** Git, GitHub, Bitbucket, Docker, Kubernetes, AWS, GCP, Azure, FastAPI, Flask

**Data Science:** Machine Learning, Deep Learning, Computer Vision, Natural Language Processing, Cloud Computing

**Agentic AI:** Tool Calling, Google ADK, A2A, LangGraph, CrewAI, Autogen, and Open AI Agentic SDK

## EXPERIENCE

### AI Engineer Intern

Sept 2023 - Apr 2024

Gujarat, India

Plutomen Technologies Pvt. Ltd.

- Architected automated parsing pipeline using **LlamaParse** and **Mistral OCR**, boosting data extraction accuracy from 45% to 89% and eliminating 15 hours/week of manual compliance review, using **LlamaIndex** and **LangChain** as orchestration
- Improved chatbot answer relevance by **30%** through **optimized chunking** and semantic search using **Qdrant**, **Pinecone**
- Designed and deployed **Flask REST APIs** backend with 8 endpoints, integrating **PostgreSQL** for chat history and enabling real-time RAG responses with less than 200ms latency for client applications
- Identified and resolved 15+ model failure modes through systematic testing, improving chatbot robustness by 12%

### Research Assistant

Apr 2023 - Jun 2023

Gujarat, India

Charotar University of Science and Technology

- Constructed a **U-Net** for MRI segmentation, achieving 91% accuracy by **training models** on preprocessed DICOM images
- Benchmarked model using IoU metrics, achieving a **15%** improvement over baseline CNN architectures
- Applied advanced AI techniques in medical imaging, demonstrating the impact of AI professionals in healthcare

### Machine Learning Engineer Intern

May 2022 - Jun 2022

Gujarat, India

NXON Pvt. Ltd.

- Designed a CNN achieving over **99%** test accuracy on MNIST by **fine-tuning** convolutional and pooling layers
- Applied dropout regularization techniques, improving model generalization by **7%** and reducing overfitting risks
- Refined architecture to boost classification reliability, streamlining deployment in production-ready AI Systems

## PROJECTS

### ClassTopper: Full Stack Multi-modal (Text + Video) RAG Platform

- Led 3-person team building multi-modal RAG chatbot for videos and documents, improving study efficiency by 40%
- Enhanced content retention by **60%** via mind-mapping feature, visualizing topic relationships for better knowledge recall
- Engineered adaptive quiz system generating personalized assessments from knowledge gap analysis, boosting test performance by 27% and identifying weak areas with 85% accuracy

### Code Review AI System with Parameter-Efficient Fine-Tuning

- Engineered AI-powered code review system using **LoRA fine-tuning** on T5 transformer, reducing trainable parameters by 99% (220M to 2M) while achieving 60% faster training convergence across **distributed multi-GPU infrastructure**
- Built distributed MLOps pipeline with **PyTorch DDP** and ClearML tracking across 60K+ training steps, reducing debugging cycles by 40% through automated checkpointing and versioning
- Achieved multi-language code intelligence across 10 languages with **CodeBLEU evaluation** (AST matching, dataflow analysis) and **beam search optimization** for ranked code suggestions

### Expense Tracker using MCP

- Deployed a custom MCP server on FastMCP Cloud, enabling MCP Client to manage finances through natural language
- Integrated Claude Desktop MCP Client, allowing users to query, summarize, and analyze expenses via conversational AI
- Leveraged a lightweight SQLite backend for fast, local, and secure storage with instant retrieval and real-time insights.

## CERTIFICATIONS AND ACHIEVEMENTS

**Google Cloud Computing** - Google Cloud Educational Badges (9 Badges)

**AWS Academy Graduate** - AWS Academy Machine Learning Foundations and AWS Academy Cloud Developing

**Data Science Club** - Led & Mentored juniors on data science, organized events, and AWS ML workshop for 100+ students.

**Hackathon Achievement** - Earned 2nd rank (Silver Medal) in the 2023 SolutionChallenge Hackathon at CHARUSAT