

# BHAVUK AGRAWAL

bhavukagrawal1409@gmail.com | +91-7060802357 | github.com/bhavuk1409 |  
linkedin.com/in/bhavukagrawal1409

## EDUCATION

**Bennett University**, Greater Noida, Uttar Pradesh, India  
Bachelor of Technology in Computer Science | CGPA: 9.04/10

Aug 2023 – 2027

## EXPERIENCE

**Machine Learning Intern** | *National University of Singapore (NUS) Singapore (On-Site)* June 2024 – July 2024

- Developed TF-IDF-based medical report analysis tool to accelerate text processing workflows for the research team
- Automated deployment infrastructure using AWS CloudFormation, reducing manual setup effort
- Deployed ML models on AWS SageMaker and optimized pipeline execution using AWS services
- Collaborated with researchers to transform academic ML/NLP concepts into functional prototypes and demos

## PROJECTS

### R-Search: Reinforcement Learning for Multi-Source Search

github.com/bhavuk1409/llm-reasoning-search

- Developed novel GRPO-based RL fine-tuning pipeline for LLM reasoning with structured output generation, integrating Exa multi-source search API and LoRA parameter-efficient training with HuggingFace and WandB support
- Designed composite reward function combining semantic coherence, DAG validation, and format compliance with curriculum learning, optimizing training stability and convergence
- Achieved 25-30% accuracy improvement on HotpotQA-hard benchmark, 40% better format compliance, and 35% VRAM reduction through LoRA optimization, enabling training on consumer GPUs

### LLM-Enhanced Trading Analysis Platform

github.com/bhavuk1409/llm-trading | llm-trading-platform.streamlit.app

- Developed multi-agent trading system with 4 specialized AI agents (Technical, Fundamental, Sentiment, Risk) using Groq LLM inference and LangChain for coordinated stock analysis and decision-making
- Integrated real-time news sentiment via Exa API with Streamlit dashboard featuring technical indicators and JSON export

### AI-Powered Real-Time Video Conferencing Platform

github.com/bhavuk1409/intellimeet\_updated

- Architected scalable platform with real-time multilingual subtitles (Whisper), auto-summarization (BART), and RAG chatbot (LLaMA + LangChain) on AWS-compatible infrastructure with WebRTC for low-latency communication

## TECHNICAL SKILLS

**Languages:** Python, C++, Java, JavaScript, SQL

**ML/DL Frameworks:** PyTorch, TensorFlow, Transformers, LLM Fine-tuning (LoRA/QLoRA/PEFT), Reinforcement Learning (PPO, GRPO)

**AI/LLM Tools:** LangChain, LangGraph, RAG, Hugging Face, Vector Databases, MCP

**Cloud & DevOps:** AWS, Docker, Git

**Web Development:** FastAPI, React, MongoDB

**Core CS:** Data Structures & Algorithms, Problem Solving

## ACHIEVEMENTS & CERTIFICATIONS

- Selected for competitive on-site ML/AI research internship at National University of Singapore (acceptance rate <5%)
- Solved 400+ Data Structures & Algorithms problems across coding platforms | **LeetCode Rating:** 1644 (Top 25%)
- Active contributor to open-source AI projects with focus on LLM reasoning and reinforcement learning applications
- NVIDIA:** Getting Started with Accelerated Computing in Modern CUDA C++
- DeepLearning.AI:** Improving Deep Neural Networks – Hyperparameter Tuning, Regularization and Optimization
- UC San Diego:** Data Structures Specialization