Arnav Gupta

+91-95185xxxxx | arnavstudies28@gmail.com | linkedin.com/in/arnav-gupta | github.com/ArnavG-728

Professional Summary

Computer Science Engineering student specializing in Artificial Intelligence (AI), Machine Learning (ML), and Cloud Computing with 3+ years hands-on experience building production applications. Experienced in developing Retrieval Augmented Generation (RAG) systems, Computer Vision models, and scalable ML pipelines using Python, JavaScript, and cloud technologies. Delivered 60-80% performance improvements in enterprise applications through optimized algorithms and system architecture. Seeking Software Engineer, Machine Learning Engineer, or Data Scientist positions.

Education

SRM Institute of Science and Technology Chennai, Tamil Nadu

Bachelor of Technology in Computer Science Engineering, Cloud Computing Specialization *Jun 2022 – May 2026*

- Current GPA: 9.09/10.0 — Relevant Coursework: Data Structures, Algorithms, Database Management Systems, Computer Networks, Operating Systems, Software Engineering

Technical Skills

Programming Languages: Python, C++, SQL, JavaScript
Machine Learning & AI: HuggingFace, OpenCV, Numpy, Pandas, TensorFlow, PyTorch, Scikit-learn
GenAI Tools: LangChain, LangGraph, LangSmith, LlamaIndex
Database Technologies: MySQL, PostgreSQL, MongoDB, Neo4j Graph Database, Google BigQuery
Cloud Platforms: AWS, Azure, GCP, Salesforce Lightning Platform
DevOps & Tools: Docker, Git, Postman
Web Technologies: FastAPI, Flask, Django, Node.js, React.js, Next.js, RESTful APIs, GraphQL
Data Engineering: ETL Pipelines, Pub/Sub, Data Warehousing, Stream Processing
Specializations: NLP, Generative AI, Deep Learning

Professional Experience

Generative AI Software Engineering Intern May 2025 – Aug 2025

Hexaware Technologies Limited *Chennai, Tamil Nadu*

- Architected CodeChat RAG (Retrieval Augmented Generation) system processing 50,000+ lines of code across 12 enterprise codebases using LangChain framework and UnixCoder embeddings for semantic code understanding
- Built Neo4j graph database infrastructure mapping repository relationships for 500+ files, enabling contextual code traversal and reducing query response time by 75% (from 8 seconds to 2 seconds)
- Accelerated developer onboarding process by 80% through intelligent semantic search capabilities, reducing average learning time from 2 weeks to 3 days for new team members
- Optimized natural language query processing achieving 92% accuracy in code explanation tasks across Python, JavaScript, and Java repositories

Key Projects

DocSimilarity AI Profile Ranking System Jul 2025 – Aug 2025

End-to-End ML Application — LangChain, Next.js, FastAPI, MongoDB *github.com/ArnavG-728/docsimilarity*

- Engineered multi-agent AI system for resume-job matching serving 200+ recruiters with 94% semantic similarity accuracy using LangChain and LangGraph orchestration frameworks
- Built scalable backend architecture with FastAPI processing 1,000+ concurrent requests and MongoDB database managing 10,000+ candidate profiles with sub-100ms query response times
- Launched production application on Render (backend) and Vercel (frontend) achieving 99.8% uptime and supporting 500+ daily active users during beta testing phase
- Reduced candidate screening time by 65% (from 15 minutes to 5 minutes per profile) enabling HR teams to evaluate 3x more candidates daily

Medical Image Classification for Breast Cancer Detection Jan 2025 – Mar 2025

Computer Vision & Explainable AI — Vision Transformer, OpenVINO *github.com/ArnavG-728/medical-vision*

- Trained Vision Transformer, Swin Transformer, and VGG models on 2,800 ultrasound images achieving best performance with Vision Transformer at 96.4% accuracy and 0.98 AUC score
- Enhanced inference speed by 2.3x (from 450ms to 195ms) using Intel OpenVINO optimization toolkit for real-time clinical deployment on edge devices
- Integrated Grad-CAM and Attention Rollout explainability techniques improving diagnostic confidence scores by 70% among 15 medical professionals during validation testing
- Processed 50GB+ medical imaging dataset with 95% data quality score through automated preprocessing pipelines using OpenCV and custom augmentation strategies

Multi-Cloud Real-Time Fraud Detection System Feb 2025 – May 2025

MLOps Pipeline — AWS, GCP, LightGBM, BigQuery, SageMaker *github.com/ArnavG-728/fraud-detection*

- Architected hybrid AWS-GCP fraud detection pipeline processing 1,200+ transactions per second with 99.7% system availability and sub-second fraud classification response times
- Generated 2.5 million synthetic transaction records using Faker library and established Google Pub/Sub streaming infrastructure handling 50MB/hour data throughput
- Orchestrated ETL pipeline using AWS Glue transforming raw financial data and feeding BigQuery data warehouse supporting 500GB+ transaction history analysis
- Benchmarked ML algorithms achieving optimal results with LightGBM (0.94 AUC), outperforming Random Forest (0.91 AUC) and Logistic Regression (0.87 AUC) across 100,000 test transactions
- Productionized real-time API using AWS Elastic Beanstalk serving 50,000+ daily predictions with SHAP explainability and automated SageMaker model retraining pipeline

Achievements and Awards

Hexaware Premier League Hackathon 2025 - First Place Winner Aug 2025

Document Similarity AI System — Prize: INR 50,000 *Chennai, Tamil Nadu*

- Secured 1st place among 150+ teams by building end-to-end document similarity system using multi-agent architecture with LangGraph, LangChain, Next.js, and MongoDB
- Presented solution to industry executives including cricket legend Rahul Dravid, demonstrating 85% improvement in document matching accuracy over baseline systems

Certifications and Training

Cloud & Platform: Salesforce Platform Developer I, Google Cloud Data Analytics Certificate, AWS Academy Machine Learning Graduate, AWS Academy Data Engineering Graduate
Technical Skills: BigQuery Insights Skill Badge (Google Cloud), Computer Architecture (NPTEL), Database Management Systems (Scaler), Computer Networks and Internet Protocol (NPTEL)
Data Science: Data Analytics with Python (NPTEL), Introduction to Data Science (Cisco), Cloud Computing Fundamentals (IBM)