

AKARSHAN GUPTA

+91 9769691096 ♦ Mumbai, MH

guptaakarshan10@gmail.com ♦ [LinkedIn](#) ♦ [GitHub](#)

OBJECTIVE

Highly motivated AI/ML Engineer seeking a full-time role to design, develop, and deploy scalable machine learning models and AI applications. Passionate about leveraging advanced algorithms and cloud solutions to solve complex real-world problems and drive innovation.

EDUCATION

Bachelor of Computer Science Engineering, A.C. Patil College of Engineering, Navi Mumbai
CGPA: 8.29/10.0

HSC Science, St. Rocks Junior College of Commerce and Science
Percentage: 65.23%

SSC, St. Xaviers High School
Percentage: 83.80%

CERTIFICATIONS

AWS Certified Solutions Architect – Associate: Demonstrated proficiency in designing scalable, highly available, fault-tolerant, and cost-effective cloud architectures on Amazon Web Services (AWS).

Theory of Computer (NPTEL): Completed a comprehensive course covering fundamental concepts in theoretical computer science, including automata theory, computability, and complexity theory.

AWS Cloud Practitioner: Validated foundational understanding of AWS cloud concepts, services, security, architecture, pricing, and support.

SKILLS

Core AI/ML: Data Structures, Algorithms, System Design, Large Language Models (LLM), Diffusion Models, Machine Learning, Deep Learning, Generative AI, Computer Vision, Natural Language Processing (NLP), Model Optimization, Predictive Modeling, Algorithm Development, MLOps

Programming Languages: Python, JavaScript, C++, MySQL, TypeScript, Node.js

Frameworks & Libraries: PyTorch, TensorFlow, React, Next.js, Dot.net, Prisma, Django, NumPy, Pandas, Scikit-Learn, OpenCV, NLTK, Gradio, Bootstrap, Vite.js, CUDA, Tailwind CSS, Hugging Face Transformers, Matplotlib, FFmpeg

Tools & Platforms: Firebase, AWS, MongoDB, Git, Apache XAMPP, Jenkins, Docker, IBM Cloud, IBM Watson, Jupyter Notebooks, Version Control (Git), PostgreSQL, Ollama, Azure API

EXPERIENCE

AIML Intern

June 10, 2024 – July 6, 2024

CsrBox

Remote, MH

- Engineered/deployed AI conversational agent for restaurant website (IBM Cloud), handling 100s daily queries, boosting engagement.
- Developed/optimized AI/ML predictive models (IBM Cloud, Watson, Jupyter), boosting accuracy to 90% in classification tasks.
- Streamlined ML evaluation/generation (IBM Cloud automation), enhancing efficiency and reducing manual processing by 30%.

- Designed/deployed potato disease classifier (Gradio UI), demonstrating end-to-end cloud deployment, processing 50 classifications/min.

AI Engineer Intern

Quantal AI

September 2025 – November 2025
M, MH

- Engineered an end-to-end AI automation system that processed 100 product images using OCR and BM25-based text matching, improving Shopify product validation accuracy by 95%.
- Built and deployed a full-stack application (React.js + FastAPI) on AWS Lightsail, integrating GCP authentication and Google Drive synchronization, achieving 99% uptime and reducing manual upload time by 60%.
- Developed and launched a real-time Voice Agent using LiveKit Cloud and ElevenLabs API, enabling users to create appointments and reminders via Google Calendar API with 95% voice command accuracy.
- Architected and optimized cross-cloud infrastructure between AWS and GCP for the wine automation system, leveraging AWS Lightsail for deployment and GCP APIs for authentication, reducing backend latency by 25% and improving overall scalability.

PROJECTS

LegalEase: AI-Powered Legal Document Processing

[\[Try it\]](#)

- Built RAG system with Groq (Llama 3.3) and ChromaDB for scalable legal contract analysis and retrieval
- Developed NLP pipelines for automated risk analysis, contract comparison, and clause simplification
- Implemented hybrid embedding strategy (Google Gemini API + local fallback) for resilient semantic search
- Reduced document processing latency by 40% using Groq's inference engine for real-time PDF summarization

Local AI Chatbot with Ollama & Gemma 2B

[\[Try it\]](#)

- Engineered full-stack, local-first ChatGPT-style AI app (Ollama, Gemma 2B LLM) for sub-second streaming.
- Developed Node.js/Express backend with persistent chat history (PostgreSQL/Prisma), storing 1000s of sessions.
- Built dynamic Next.js/React frontend (Tailwind CSS) with real-time chat, auto-titling, multi-session management.
- Architected fully local AI solution, supporting multiple LLMs without external APIs.

SmartCity Studio: AI-Powered Urban Planning with 3D Visualization

[\[Try it\]](#)

- Built 3D smart-city platform with React, Three.js, and Leaflet featuring dual-view toggle and 8×8 grid-based planning
- Engineered AI-powered infrastructure designer with real-time suggestions via OpenAI/Gemini APIs
- Developed predictive analytics dashboard with climate risk assessment, demographic simulation, and exportable PDF reports
- Implemented layer-based visualization for zoning, utilities, and green spaces with GLB/JSON export functionality

Food Classifier: Deep Learning for Image Recognition

[\[Try it\]](#)

- Developed robust food classification deep learning model (Food-101: 5-6 GB, 101 categories, 101K images), resolving accuracy/overfitting/underfitting issues.
- Implemented/trained custom mini-VGG/mini-ResNet for experimental design, achieving 85% accuracy (10% over baseline).
- Utilized GPU acceleration to reduce training time (days to hours); deployed Gradio web interface (pre-trained ResNet-50) for real-time predictions (<0.5s/image).
- Implemented project using Python/PyTorch for comprehensive development, facilitating ~100 user interactions/full-stack ML deployment.