Kunal Mohare

linkedin.com/in/kunal-mohare github.com/Mohare786-Kunal

kunaldmohare@gmail.com +91 9156495035

EDUCATION

RAMDEOBABA UNIVERSITY Nagpur, Maharashtra
*B.Tech in Computer Science (Exp. **2026**)* - **GPA: 7.5/10** **2022** – Present
BHAVAN'S BHAGWANDAS PUROHIT VIDYA MANDIR Nagpur, Maharashtra
*CBSE 10th Board – **92.8%*** **2006 - 2022**

EXPERIENCE

BINGECLIP AI – *AI/ML Engineer* January **2025** – Present
Bangalore, Karnataka

- Architected scalable, production-ready ML systems on **GCP** integrating open-source models (e.g., Hugging Face Transformers, BERT), reducing processing time by **25**% via optimized data pipelines.
- Led a team of **3** interns in developing Multi-Agent AI solutions using **CrewAI, LangGraph** and **TensorFlow**, processing over **1,000** documents monthly with **95**% accuracy through advanced multi-threaded processing.

SENSLYZE – *AI Developer* September – November **2024**
Nagpur, Maharashtra

- Launched an MVP for YES BANK with **RAG** pipelines using **LangChain** and **FAISS**, enhancing workflow efficiency by **40**% with semantic search and real-time retrieval, ensured **99.9% reliability** with robust error handling and **AWS**-based load balancing, and implemented **guardrails** with rule-based filters to prevent hallucinations and maintain data integrity.
- Deployed Agentic AI systems with **LangGraph**, improving processing efficiency by **30**% via dynamic coordination, reduced downtime by **95%** with **Redis** caching and failover strategies, and optimized reliability with **CloudWatch** monitoring and self-healing mechanisms, handling 10,000+ concurrent requests seamlessly.

PROJECTS

Multi-Agent AI System — DEMO : [**YouTube**] GitHub

- Engineered a Multi-Agent AI platform integrating **Google Workspace**, **GitHub**, and **LinkedIn APIs** with **Redis** caching, achieving real-time data synchronization.
- Automated OS-level tasks using **AppleScripts** and **PowerShell**, boosting productivity by **35**% with custom task schedulers.

Multimodal - Retrieval Augmented Generation (RAG) System GitHub

- Built a multimodal **RAG** system with implementation of the **Model Context Protocol (MCP)** using **CLINE plugins** to manage and optimize context-aware retrieval, ensuring efficient handling of large-scale contexts by structuring the thesis of retrieved data into hierarchical embeddings; this enabled dynamic context pruning, improved relevance scoring by **20**%, and supported scalability for datasets exceeding **1 lakh entries**.
- Deployed a **FastAPI** backend with **Redis** caching and **LangChain**, scaling **LLM** (**OpenAI**) workflows with dynamic batching for efficient multimodal data processing.

Ads CTR Optimization GitHub

- Designed **Reinforcement Learning** algorithms - **Deep Q-Learning, Upper Confidence Bound (UCB), and Thompson Sampling** to optimize ad placements, increasing **CTR** by **25**% on a **50,000**-user dataset.
- Analyzed large-scale data patterns with **Pandas** and **TensorFlow** for dynamic, real-time ad adjustments.

Road Object Detection GitHub

- Designed a **CNN**-based system with **OpenCV** and **YOLOv5** for real-time detection, achieving **90**% accuracy on a **1,000**-frame dataset.
- Optimized for low-latency video analysis with multi-threading, targeting autonomous vehicle applications, reduced inference time by **40**% using GPU acceleration with **CUDA**, and implemented a batch processing pipeline to handle varying frame rates, enhancing scalability for deployment in resource-constrained edge devices.

VideoInsight Pro GitHub

- Engineered a cloud-based SaaS platform that converts YouTube videos into concise summaries, detailed transcripts, and trains a custom AI assistant tailored to video content, leveraging **AWS S3** for storage, **FastAPI** for API deployment, **Open AI API**, **Anthropic Claude**, and **Gemini LLM** for NLP, and **PyTorch** for model training.
- Achieved a **40%** user retention rate through iterative feedback and feature enhancements by college students, with an average processing time reduction of **30%** using optimized batch processing, and integrated **Docker** for containerization and **Redis** for caching, ensuring scalability and reliability for high-traffic scenarios.

ACHIEVEMENTS

- Secured **First Place** in Groclake Agentic AI Hackathon, leading the team with an innovative AI solution.
- Secured **top 5 Finalists** position in Eyantra IITB Competition, showcasing advanced DL/ML and robotics integration.
- Successfully **Integrated the Agentic AI System for Google Workspace** into my university's ecosystem, enhancing productivity for teachers across departments.
- Solved **250+ LeetCode** problems, demonstrating strong problem-solving and algorithmic expertise.
- Published technical blogs on **Hashnode** with **1K+ views**, establishing a presence in the tech community.

SKILLS

- **Programming:** Python, Java
- **AI/ML:** LangChain, Hugging Face, TensorFlow, PyTorch, OpenCV, Reinforcement Learning, Scikit-learn
- **Development:** FastAPI, Node.js, Express.js, Docker, Kubernetes, GitHub Actions
- **Databases & Tools:** Git, AWS, GCP, ChromaDB, FAISS, MongoDB, PostgreSQL, Redis
- **Concepts:** System Design, Algorithms, Data Structures, Distributed Systems, Cloud Computing
- **Soft Skills:** Leadership, Adaptability, Teamwork, Problem-Solving, Communication, Time Management