

KARTAVYA BALUJA

07988467579 kartavyabaluja453@gmail.com [linkedin.com/in/kartavya-baluja-214ba1256](https://www.linkedin.com/in/kartavya-baluja-214ba1256) github.com/kartavya4874

Professional Summary

Final-year Computer Science student and practicing AI Engineer with hands-on experience building LLM-powered applications, RAG workflows, and Agentic AI systems. Proven ability in end-to-end ML pipeline development, data preprocessing, and deploying AI solutions via REST APIs. Certified in Azure AI, Oracle AI Vector Search, and multiple AI platforms, with a published research paper in interpretable machine learning.

Education

Geeta University

Bachelor of Technology in Computer Science

Expected: August 2026

India

Technical Skills

Core Python Skills: Python, Pandas, NumPy, scikit-learn, TensorFlow, Keras

AI/ML Technologies: Large Language Models (LLMs), RAG (Retrieval-Augmented Generation), Agentic AI, NLP, Deep Learning

Web Frameworks: FastAPI, Flask (Django knowledge)

Data & Visualization: Data Cleaning, Data Pipelines, Matplotlib, Seaborn, Statistical Analysis

Databases & APIs: Vector Databases (Oracle AI Vector Search), Neo4j, REST APIs, SQL

Cloud Platforms: Azure AI, Oracle Cloud Infrastructure (OCI), Basic AWS/GCP knowledge

AI Platforms: Hugging Face, IBM Watson, Streamlit

Development Tools: Git, GitHub, Jupyter Notebooks, Google Colab, Linux

Experience

AI Engineer

January 2026 – Present

Geeta Technical Hub

India

- Design and develop AI-powered applications and tools leveraging state-of-the-art LLMs and deep learning frameworks
- Build and maintain end-to-end ML pipelines including data ingestion, preprocessing, model training, and deployment
- Implement RAG (Retrieval-Augmented Generation) and Agentic AI workflows to enhance application intelligence
- Collaborate with academic and technical teams to research and apply emerging AI methodologies
- Optimize model performance through experimentation, hyperparameter tuning, and evaluation frameworks

ML & AI Intern

June 2025 – January 2026

White And Box – Tech Products & Services

Bengaluru, Karnataka, India

- Build and deploy LLM-powered AI solutions using Python frameworks, focusing on practical business applications
- Develop data pipelines for ML model training and inference using Pandas, NumPy, and TensorFlow
- Collaborate with cross-functional teams to integrate AI features into production environments via REST APIs
- Research and evaluate emerging AI trends including RAG workflows and Agentic AI implementations
- Clean and prepare large datasets for machine learning tasks, ensuring data quality and model performance

Projects

AI Agent Platform – Multi-Modal LLM Application | Python, FastAPI, Streamlit, LLMs

2025

- Built production-ready AI platform integrating multiple Large Language Models for writing, coding, and content generation
- Developed FastAPI backend with REST API endpoints for seamless frontend-backend communication
- Implemented data preprocessing and model orchestration workflows using Pandas and NumPy
- Created interactive Streamlit frontend enabling real-time AI interactions and visualizations

Deep Learning Food Recognition System | Python, TensorFlow, Keras, CNN

2024

- Preprocessed and cleaned Food-101 dataset using NumPy and Pandas for deep learning training
- Built CNN architecture and fine-tuned InceptionV3 model using TensorFlow and Keras
- Implemented data pipelines for model training, validation, and inference workflows
- Achieved high accuracy through hyperparameter optimization and performance analysis

Aashiyana – AI Chatroom with NLP Moderation | Python, NLP, FastAPI, Git

2025

- Developed real-time chat application using Python and FastAPI with integrated NLP-based content filtering
- Built data processing pipelines for message analysis and intelligent moderation using ML algorithms
- Implemented RESTful APIs for authentication, messaging, and AI-driven security features
- Managed version control and collaboration using Git and GitHub throughout development lifecycle

Publications & Research

Bridging the Accuracy-Explainability Gap: A Survey of Interpretable ML Techniques

2024

- Published in Anveshan: Multidisciplinary Journal of Geeta University, Volume 2, Issue 1, pp. 13–26
- Conducted comprehensive research on interpretable ML techniques and their trade-offs with model accuracy
- Analyzed emerging approaches for bridging the gap between black-box performance and explainability

Certifications

Microsoft Certified: Azure AI Engineer Associate – Microsoft (2025)

Oracle AI Vector Search Certified Professional – Oracle (2025)

Oracle Cloud Infrastructure 2023 AI Certified Foundations Associate – Oracle (2024)

Salesforce Certified AI Associate – Salesforce (2024)

Neo4j Certified Professional – Neo4j (2024)

Artificial Intelligence Primer Certification – Infosys Springboard (2024)