

RAHUL SHARMA

+91 7349485773 — [✉ rahulsharmavishwakarma@gmail.com](mailto:rahulsharmavishwakarma@gmail.com) — [🌐 linkedin/in/rahulsharmavishwakarma](https://www.linkedin.com/in/rahulsharmavishwakarma)
[🐙 github/rahulsharmavishwakarma](https://github.com/rahulsharmavishwakarma) — [🌐 rahulsharmavishwakarma.github.io](https://rahulsharmavishwakarma.github.io)

Summary — Recent Computer Science & Engineering graduate with a strong foundation in AI, machine learning, and software development. Experienced in developing and fine-tuning AI models for computer vision, NLP. Proven track record in research and practical implementations, seeking full-time opportunities in Software Engineering, AI and machine learning.

Skills

Programming Languages: Python, SQL, JavaScript, C++ , Java, HTML, CSS

Libraries: Numpy, Pandas, Matplotlib, Seaborn, TensorFlow, PyTorch, Scikit-Learn, Django, FastAPI, NLTK, Langchain, MLflow

Tools: Azure, GCP, AWS, Git/GitHub, Docker, Linux, VS Code, HuggingFace, Jupyter, Hadoop, Data Structures

Databases: MySQL, SQLite, MongoDB, Neo4j, Vector Databases, Pinecone, Milvus

Machine Learning: Deep Learning, Computer Vision, NLP, GANs, LLMs, Transformers, VAEs, Diffusion Models, RAGs, Generative Models, Mutlimodal Models

Experience

Zysec AI

Oct 2024 – Present

Associate Generative AI Engineer

- Developed Retrieval-Augmented Generation (RAG) models to improve information retrieval and response accuracy for cybersecurity applications.
- Utilized LangChain to build agentic applications that support real-time decision-making tailored for cybersecurity needs.
- Optimized large language models (LLMs) to manage memory constraints effectively, enhancing scalability and performance for complex, context-heavy tasks.
- Integrated Neo4j for graph retrieval to improve context handling in LLMs, allowing for enhanced understanding and response in multi-turn dialogues.

HacktivSpace

Sept 2024 – Present

Machine Learning Engineer

- Working on Retrieval-Augmented Generation (RAG) for context retrieval in EdTech applications, utilizing Milvus and Neo4j to manage and retrieve data effectively.
- Developing conversational memory in LLM applications to maintain context across sessions, improving engagement and continuity in AI interactions.
- Using frameworks such as LangChain and CrewAI to create AI agentic workflows for various tasks, supporting automation and enhanced responsiveness.

HKBK College of Engineering

Aug 2023 – May 2024

Undergraduate Researcher

- Developed a model to analyze and respond to questions on biomedical images with over 70% accuracy.
- Conducted research on fine-tuning large language models like LLaMA for medical question-answering.
- Designed an R-CNN-based model to extract templates and information from documents for fraud analysis.

Varcons Technologies

Aug 2023 – Sep 2023

Machine Learning Intern

- Built a predictive sentiment analysis model for stock price prediction, achieving over 80% accuracy.
- Conducted research on various machine learning topics in computer vision and language processing systems.

Publications

Beyond Imagery: AI-Enhanced Diagnostic Assistant for Cancer and Tumor Diagnosis using Radiology Imaging [link](#)

Authors: Dr. Nandha Gopal S M, Rahul Sharma, Nithin M, Prajwal B R, Prashanth Kalgonda

Mar 2024

International Journal On Engineering Technology and Sciences (IJETS)

Education

Visvesvaraya Technological University

Bengaluru, India

Bachelor of Engineering(B.E/B.Tech) in Computer Science and Engineering

Sep 2020 – June 2024

CGPA: 8.6/10

Mahesh PU College

Bengaluru, India

12th Standard/PU in PCME

May 2019 – Mar 2020

Percentage: 78% — Marks: 463/600

Projects

- Medi-Care: VQA for Medical Imaging** [link](#)
- Developed a multimodal model for analyzing and responding to medical images using a diverse VQA dataset.
 - Fine-tuned LLaMA models for question answering, achieving over 75% accuracy.
 - Incorporated RAG techniques to enhance model accuracy and credibility for knowledge-intensive tasks.
- Resume Retrieval and Question-Answering System** [link](#)
- Developed an AI-powered resume retrieval system using Milvus vector database for efficient storage and retrieval of resume embeddings, enabling hybrid search with sparse and dense vector representations..
 - Integrated LLMs for generating human-like responses to queries, leveraging advanced NLP libraries like PyMilvus and Hugging Face's InferenceClient for seamless interaction between vector database and language model.
- Diffusion Detect** [link](#)
- Integrated Stable Diffusion and YOLO for text-to-image generation and object detection.
 - Implemented a pipeline for text prompts, image generation via Stable Diffusion, and YOLO object detection.
- Training YOLO object detection models for custom datasets** [link](#)
- Developed the algorithm to convert the annotations from Pascal VOC format to YOLO format.
 - Developed a YOLO-based model for detecting personal protective equipment (PPE) in images, with a focus on improving workplace safety.
 - Trained the model on a custom dataset to accurately identify various PPE items like helmets, gloves, and safety vests.

Certifications

- MongoDB Certified Associate Developer** [link](#)
July 2024
MongoDB
- Oracle Cloud Infrastructure 2024 Generative AI Certified Professional** [link](#)
June 2024
Oracle
- Microsoft Certified: Azure Data Scientist Associate** [link](#)
June 2024
Microsoft
- Deep Learning Specialization** [link](#)
Aug 2023
DeepLearning.AI
- Microsoft Certified: Azure AI Fundamentals** [link](#)
Feb 2023
Microsoft