PRASAD P R

+91 9740146523 | shenoy.prasad20@gmail.com | Bengaluru, India

LinkedIn: https://www.linkedin.com/in/prasad-pr/ | GitHub: https://github.com/prp20

PROFESSIONAL SUMMARY

Results-driven Senior AI / ML Engineer with 7+ years of experience in designing, developing, and optimizing AI/ML systems for production environments. Specialized in Generative AI, LLMs, RAG pipelines, NLP, Computer Vision, and AI Infrastructure Optimization. Proven track record in delivering enterprise-grade, scalable, and efficient AI/ML solutions using Python, TensorFlow, PyTorch, Hugging Face, OpenVINO, ONNX, and Kubernetes. Adept at building AI platforms from experimentation to deployment with a focus on performance, scalability, and automation.

PROFESSIONAL EXPERIENCE

AI Software Development Engineer

Intel Corporation - Bengaluru, India | Jun 2022 - Present

Technologies: Python, PyTorch, TensorFlow, OpenVINO, ONNX, Docker, Kubernetes, LLM APIs, LangChain, LlamaIndex, NLP, HuggingFace

- Spearheaded model optimization and benchmarking for Intel GPUs, achieving 10x performance improvements for transformer-based models (Whisper, M2M100).

- Developed and deployed automated pipelines for AI benchmarking across distributed systems, reducing benchmarking time from 5 hours to 30 minutes per machine.

- Enhanced AI performance via ONNX Runtime and OpenVINO toolchain optimizations, reducing analysis time by 50%.

- Built a real-time 3D camera calibration pipeline leveraging Python, PyTorch, Open3D, and Docker, achieving 80% reduction in inference time.

- Delivered scalable AI solutions integrated with IoT devices (MQTT, video feeds) for real-time computer vision and data visualization.

- Integrated AI telemetry pipelines and performance dashboards to improve system monitoring and SLA/SLO adherence.

System Software Development Engineer

Intel Corporation - Bengaluru, India | Aug 2018 - Jun 2022

Technologies: MERN Stack, Python, Microservices, CI/CD, Docker, Kubernetes

- Led the architecture and development of a MERN-based AI Solution Delivery Platform supporting 1000+ downloads, improving access latency by 30% through microservices adoption.

- Developed internal management portals integrated with CI/CD pipelines (Git, Jenkins), reducing operational overhead by 80%.

- Delivered scalable backend services aligned with best practices in microservices, containerization, and DevOps.

EDUCATION

Executive Post Graduate Certification in AI & ML

IIIT Bangalore & upGrad | Bengaluru, India | Feb 2024 - Apr 2025

CGPA: 3.63 / 4.0

Core Topics: Generative AI, LLMs, RAG, LangChain, Transformers, Prompt Engineering, Cloud AI Deployment

Bachelor of Engineering (B.E.), Computer Science & Engineering

JSS Academy of Technical Education | Bengaluru, India | Aug 2014 - Jul 2018

Graduation: Distinction

KEY PROJECTS

Generative AI for Automated Reports (Internal Tool)

- Designed a Generative AI system leveraging LLMs and Prompt Engineering to automate report collation from distributed email inputs.

- Reduced reporting efforts from 5 days to 30 minutes through AI-driven data extraction and summarization.

AI-Based Resume Screening (Internal Tool)

- Built an LLM-powered platform using LangChain and embeddings for semantic resume-job description matching.

- Reduced manual screening time from hours to minutes, enhancing recruitment efficiency.

AI-Driven Financial Analysis Platform

- Developed an Agentic AI system using Python, Phidata, and Pandas for automated company financial analysis.

- Automated data gathering and analysis workflows for investment decision support.

SKILLS

Programming Languages: Python, JavaScript

AI / ML Techniques: Deep Learning, Generative AI, LLMs, NLP, Computer Vision, RAG, Knowledge Graphs, AI Agents

Frameworks / Libraries: PyTorch, TensorFlow, Hugging Face, OpenVINO, ONNX, LangChain, LlamaIndex

AI Infrastructure: Docker, Kubernetes, Microservices, REST APIs, CI/CD (Jenkins, GitHub Actions)

Data Processing / Visualization: Pandas, NumPy, Open3D, SQL, NoSQL, Data Visualization, Quantitative Analysis

Tools: Git, Jupyter, VS Code, Postman, Prometheus, Grafana

CERTIFICATIONS

- Deep Learning with PyTorch for Image Segmentation - Coursera

- LangChain for LLM Applications - Udemy

- Generative AI with Large Language Models - Coursera

AWARDS & RECOGNITION

- Divisional Recognition Award (DRA) - Intel

- Innovation Star Award - Intel

LINKS

GitHub: https://github.com/prp20

LinkedIn: https://www.linkedin.com/in/prasad-pr/

AI Financial Analysis Project: https://github.com/prp20/Fundamental_stock_analysis_using_phidata

Coursera: Deep Learning with PyTorch: https://coursera.org/share/23d1f90b78230464760aa1d075c626ab

Udemy: LangChain for LLM: https://www.udemy.com/certificate/UC-0b0ea886-e7d3-4bd7-91cd-741d4399b56f/

Coursera: Generative AI LLMs: https://coursera.org/share/383ea7095df5000ac6c46acc30b037a5