

Puvith Kumar

Software Engineer - AI | Python | Production RAG, LangGraph & Backend Systems

puvithkumar2004@gmail.com | 935 324 8764 | portfolio | LinkedIn | GitHub | India

AI Engineer experienced in production RAG, agentic AI, LangGraph, LLM deployment, PostgreSQL, Redis, and LLM APIs. Currently building AI workflows at Sky360.ai and developing GraphBash, an MCP-based system agent.

Education

SAHYADRI COLLEGE OF ENGINEERING & MANAGEMENT

Nov 2022 – June 2026

Bachelor of Engineering in Computer Science (CGPA 8.83)

Experience

SKY360.ai - Software Engineer - AI | Onsite July 2026 - Present

- Developed production-oriented RAG applications for structured and unstructured company data using LangGraph, LLMs, vector databases, PostgreSQL, Redis, and SQL. Built agentic workflows for query rewriting, routing, context management, tool execution, response generation, and validation.
- Developed document and tabular question-answering systems achieving a 70% improvement in response accuracy while enhancing memory management, prompt quality, and observability.

ENDAVA - Software Engineer Intern | Remote Jan 2026 – May 2026

- Worked on frontend design, backend programming, and database management. Built and maintained responsive web pages and implemented database operations using SQL while following software development best practices.

TECHNICAL CAREER EDUCATION (TCE) - ML Intern | Mangalore. Feb 2025 – Oct 2025

- Developed a gender classification system for smart toilets and a classroom monitoring solution using YOLO based object detection, both deployed on Raspberry Pi for real-time, edge-based automation.

ERICSSON - Intern | Remote. Sep 2024 – Oct 2024

- Developed a POC for the "E2E Development Pipeline Exploration," integrating AI and ML into the SDLC to automate the generation of SRS, HLD, LLD, and other software documentation, reducing manual effort.

Projects

GraphBash – MCP and LangGraph Linux System Agent GitHub

- Designing an agentic AI system (LangGraph + MCP) for securely managing Ubuntu machines from a centralized control plane, with policy-based tool selection and human-approval gates for system-changing operations. (In progress)
- Integrated Model Context Protocol (MCP) with policy-based tool selection, risk classification, and human approval gates to prevent unrestricted system access and enable controlled execution of system-changing operations.

Tools Used: Python, FastAPI, LangGraph, Model Context Protocol (MCP), Pydantic, PostgreSQL, Redis, Docker, Linux, SSH, Pytest.

SareeFusion – AI-Powered Saree Design Automation Platform. GitHub

- Developed a smart saree design platform enabling users to combine pallu, border, and body elements, automatically generating weaving graphs for handloom production. Used YOLO trained via Roboflow for precise detection of design elements; explored multiple diffusion models, with Gemini API used for final image generation.

Tools Used: React, Flask, YOLO, Roboflow, Gemini API, OpenCV, Pillow, NumPy, REST API, Diffusion Models

Deep Research Agent (Agentic AI) GitHub

- Developed a multi-agent deep research system using LangGraph with Planner, Researcher, Writer, and Evaluator agents for autonomous information gathering and report generation.

Tools Used: Python, LangGraph, LangChain, Gemini, Tavily Search API, RAG, Git, GitHub.

Leadership and Volunteering

Mentor for internship at Coding Key – Guided students in project development. 2024

SOSC - AI and ML Student Expert – Conducted workshops and mentoring sessions. 2024

Certifications

Machine Learning Specialization - DeepLearning.AI Nov 19 , 2024

Skills & Technologies

AI/LLM: Agentic AI, LLMs, RAG, LangGraph, LangChain, MCP, Prompt Engineering, LLM Evaluation, NLP

LLM Infrastructure: vLLM, Hugging Face, LLM Serving, Model Quantization, OpenAI-Compatible APIs

Backend & Data: Python, FastAPI, Flask, PostgreSQL, Redis, MongoDB, MySQL, REST APIs

ML/CV: PyTorch, TensorFlow, scikit-learn, YOLO, OpenCV, NumPy, Pandas

Development: Java, JavaScript, C, React, Spring Boot, Node.js, Docker, Git, Linux