PRASANTH MAMIDISETTI

AI/ML Engineer | Generative AI · Large Language Models · RAG · AI Agents

North Richland Hills, TX | +1 469-839-5534 | mamidisettiprasanth36@gmail.com |
linkedin.com/in/mamidisetti-prasanth-b1aa92225 | github.com/prasanth3021 | prasanth3021.github.io

PROFESSIONAL SUMMARY

AI/ML Engineer with 4+ years of experience designing, building, and deploying production AI systems across Generative AI, Large Language Models, Retrieval-Augmented Generation (RAG), and autonomous AI agents. Deep expertise in Python, PyTorch, scikit-learn, LangChain, and LangGraph, with a record of measurable results: lifted classification F1 from 0.74 to 0.86 and shipped RAG pipelines and agent workflows over 50,000+ documents. Strong full-stack AI engineering foundation spanning FastAPI microservices, AWS cloud deployment (S3, Lambda, ECS), Docker, CI/CD, and SQL data platforms. Holder of three industry certifications from Oracle and Microsoft in AI foundations and agentic AI. Graduate candidate in Artificial Intelligence (December 2027), focused on building intelligent, scalable products that solve real business problems.

TECHNICAL SKILLS

AI/ML & LLMs: Generative AI, Large Language Models (LLMs), RAG Pipelines, AI Agents, Agentic AI, LangChain, LangGraph, PyTorch, scikit-learn, TensorFlow, Hugging Face Transformers, NLP, Deep Learning, Prompt Engineering, Embeddings, Vector Databases (Pinecone, FAISS), Fine-Tuning, LLM Evaluation, Semantic Search

Data & Analytics: Python, SQL (PostgreSQL, MySQL), Pandas, NumPy, Data Pipelines, ETL/ELT, Exploratory Data Analysis, Feature Engineering, Statistical Analysis, Data Validation, Data Quality, Cross-Validation, Hyperparameter Tuning, A/B Testing

Engineering & Cloud: AWS (S3, Lambda, ECS, SageMaker), Docker, Kubernetes, FastAPI, REST APIs, Spring Boot, MongoDB, Microservices, CI/CD Pipelines, Git, GitHub, Linux, API Testing

Languages: Python (primary), Java, C, C++, JavaScript, SQL, Bash

Practices: Agile / Scrum, Zoho Sprints, Sprint Planning, Code Review, Technical Documentation, Test Planning

PROFESSIONAL EXPERIENCE

BTG Holdings — Software Engineer Intern (Remote) | Aug 2026 – Present

- Build AI-driven product intelligence features using Python, AI/ML workflows, and MongoDB, improving how product data is organized, captured, and matched across search and ranking.
- Analyze structured product data, attributes, reviews, and merchant-sourced information to support product discovery and consumer-focused insights.
- Evaluate search, ranking, matching, and recommendation outputs through product benchmarking and comparative analysis, identifying data-quality issues and improvement opportunities.
- Analyze product-preference and buying-habit signals to support consumer insights and personalized product recommendations.
- Validate product records with systematic data-quality checks, improving the accuracy, consistency, and reliability of product information across the catalog.
- Test and validate APIs to verify data flow, application behavior, and integration functionality across services.
- Research emerging AI and automation approaches and contribute to AI/ML automation workflows for product intelligence and consumer-facing applications.
- Collaborate with engineering and cross-functional teams in Agile sprints (Zoho Sprints), mapping business and technical workflows, data flow, system dependencies, and integration points.

Texas Review Pvt Ltd (EdTech AI Platform) — Artificial Intelligence Engineer | Jan 2023 – Dec 2025

- Lifted classification F1-score from 0.74 to 0.86 on held-out test sets by building and evaluating deep learning and tabular ML models (PyTorch, scikit-learn) with feature engineering, hyperparameter tuning, cross-validation, and A/B testing against strong baselines.

- Delivered production RAG pipelines and LangChain/LangGraph agents over 50,000+ documents; selected embeddings, fine-tuned transformers, and LLM APIs by benchmarking on noisy real-world data.
- Integrated OpenAI GPT models into enterprise applications for intelligent document analysis and conversational AI (chatbot) applications used in production workflows.
- Built NLP pipelines for document classification, text preprocessing, sentiment analysis, and information extraction across large document collections.
- Deployed containerized AI microservices on AWS (S3, Lambda, ECS) via Docker and CI/CD; built SQL/Python data-validation checks; presented findings and recommendations to the team.
- Developed AI-powered applications and REST APIs using Python and FastAPI, delivering intelligent document processing and enterprise automation features.
- Implemented semantic search and document retrieval with vector databases (Pinecone, FAISS), improving contextual search and question-answering quality.
- Applied prompt engineering and systematic model evaluation to improve the response quality and reliability of LLM-powered features.
- Collaborated with product managers, data engineers, and QA teams in Agile sprint cycles to deliver AI features on schedule.

Infosys Limited — System Engineer Trainee | Jul 2022 – Dec 2022

- Built ML pipelines with Python and scikit-learn surfacing anomalies in noisy operational data.
- Developed Machine Learning models and NLP-based solutions for automated ticket classification and intelligent routing.
- Performed data preprocessing, feature engineering, model training, and performance evaluation on structured and unstructured datasets.
- Developed automation scripts to streamline repetitive business processes for operations teams.
- Built REST APIs integrating AI models with enterprise applications; used SQL databases for data extraction and analysis.
- Participated in code reviews, testing, and debugging within Agile software development teams.

PROJECTS

Career Copilot — MCP Server for AI-Assisted Job Search (Amazon Developer Hackathon) | 2026

- Designed and built Career Copilot, an MCP (Model Context Protocol) server in Python with FastMCP that connects AI assistants to a structured personal job-search workflow.
- Implemented three tools: a SQLite-backed job-application tracker, a resume-vs-job-description keyword matcher, and an interview quiz generator.
- Applied production GenAI engineering patterns: tool design, structured data access, and prompt-driven generation with clean, testable interfaces.
- Shipped end to end for the Amazon Developer Hackathon (October 2026): public MIT-licensed GitHub repository, live demo video, and Devpost submission.

CERTIFICATIONS

Oracle Cloud Infrastructure Certified AI Foundations Associate — Oracle (Oct 2026, score 98%)
Foundations of Agentic AI in GitHub — Microsoft (Sep 2026)
Oracle Agentic AI Certified Foundations Associate — Oracle (Sep 2026)

EDUCATION

Indiana Wesleyan University, USA — M.S. in Artificial Intelligence, Expected December 2027 (GPA 3.9)

Coursework: Machine Learning, Deep Learning, NLP, Generative AI, Data Mining, Statistics

Godavari Institute of Engineering and Technology, India — B.Tech, Civil Engineering, 2018–2021 (GPA 9.7)