PAUL ALEX SAMUEL JOHN

AI Engineer | B.Tech in CSE (AI & ML)
Bengaluru, India | paul.alexsamuel17@gmail.com | +91 7093317743
LinkedIn: https://www.linkedin.com/in/paul-alex-samuel-pasala-735a9526b

CAREER OBJECTIVE

AI Engineer with a strong foundation in Generative AI, cloud-native ML solutions, and full-stack development. Passionate about building intelligent systems that solve real-world problems using LLMs, Azure AI services, and autonomous agents. Seeking to contribute to innovative projects at the intersection of AI, cloud, and user-focused product development, while continuously advancing skills in cutting-edge technologies.

WORK EXPERIENCE

Sereno, Mumbai — AI Engineer (Full-Time) (January 2025 – December 2025)
- Built and deployed AI-driven solutions for enterprise workflows with a focus on agentic AI systems.
- Designed and integrated RAG-based multi-agent pipelines into client products, cutting manual query handling by40%.
- Optimized Azure AI Search + SQL integration, enabling faster query resolution and improved analytics.
- Mentored interns on LLM-based system design and best practices in deploying cloud-scale AI.

NxtGen Cloud Services — AI Engineer Intern (July 2024 – December 2024)
- Developed and deployed AI models for cloud-native applications, improving efficiency by 30%.
- Designed API-driven ML pipelines using FastAPI, reducing latency by 40%.
- Automated model monitoring and versioning in production.
- Worked with Docker, Kubernetes, SageMaker, Azure, and Vertex AI for end-to-end ML lifecycle.
- Built document AI pipelines using Azure Document Intelligence and OpenAI embeddings; integrated with AzureCognitive Search.

PROJECTS

Multi-Agent Text-to-SQL System (Loyalty Program) – Engineered a 6-agent framework converting NL queries toSQL; integrated with Azure SQL, Azure AI Search, Cosmos DB.
Azure Document Intelligence OCR + Embedding Pipeline – Built GenAI pipeline to extract text from OCR PDFs,implemented embeddings, and indexed in Azure Cognitive Search.
AI Agentic System with Semantic Kernel (Chinook DB Use Case) – Integrated Hugging Face LLM with SemanticKernel for querying relational tables.
ASR to ASR – Converted Kannada audio to Hindi audio by breaking into 3 parts using ASR , LLM and STT by using open-source models from Huggingface with better Prompting for Translation issues.
Autonomous RAG System with Memory & Agents – Combined LangChain Memory, GPT-4, ChromaDB withagentbased architecture, achieving 90% context retention.

SKILLS

Languages: Python, Java, SQL, C
Web & App Dev: React.js, Node.js
Frameworks: TensorFlow, PyTorch, Hugging Face, Keras, Scikit-Learn, Langgraph, MCP
Cloud Platforms: AWS, Azure, GCP
Deployment: FastAPI, Flask, Docker, Kubernetes, CI/CD
Data & Embeddings: FAISS, ChromaDB, LangChain, Pinecone
Vision & NLP: OpenCV, YOLO, R-CNN, GPT-4, BERT
Tools: Git, Jupyter, Streamlit, Google Colab
Soft Skills: Problem-solving, Teamwork, Analytical Thinking

EDUCATION

B.Tech in CSE (AI & ML) – Avanthi St. Theresa College of Engineering & Technology (2020 – 2024) | GPA: 70%
Intermediate – TP Junior College (2018–2020) | GPA: 70%

10th – Seventh Day Adventist School (2018) | GPA: 97%

CERTIFICATIONS

- AWS Certified Machine Learning
- Microsoft Azure AI Fundamentals
- Google Cloud AI Engineer
- Hugging Face Agent Tools
- Robo Coupler Certification

CONTRIBUTIONS & BLOGS

- Contributed to open-source GenAI projects.
- Wrote blogs on GenAI & ML on GitHub and Medium.
- Participated in AI research competitions and hackathons.