Alwin Paul

Cottbus – Germany
+49 176 77354827 • alwin.paulpv@gmail.com • alwinpaul.me
alwin-paul • alwinpaul1

Objective

AI Engineer with hands-on experience in **end-to-end RAG pipeline development, LLM evaluation, and prompt engineering**. Proven track record in designing **Graph RAG systems, optimizing chunking strategies, and implementing vector database solutions** for production environments. Specialized in **LLM application deployment, evaluation frameworks, and cross-functional AI integration**.

Professional Experience

Perinet GmbH **Cottbus, Germany**
Working Student - AI & Systems Engineer *07/2024 – present*

- Designed and deployed an **end-to-end Graph RAG pipeline** from scratch, implementing comprehensive data parsing, ingestion workflows, and optimized chunking strategies to process fragmented documentation across multiple sources.
- Engineered **prompt engineering workflows** and evaluated retrieval quality using multiple metrics including correctness, relevance scoring, and context precision to minimize hallucination in LLM responses.
- Implemented production-grade **vector database solutions using ChromaDB** integrated with embedding models (SentenceTransformers) for efficient semantic search and retrieval-augmented generation.
- Utilized **RAGAS evaluation framework** to rigorously assess RAG pipeline performance across faithfulness, answer relevancy, context precision, context recall, and response latency metrics for continuous optimization.
- Built knowledge graph structures using **Neo4j** to capture document relationships and enable context-aware information retrieval with improved semantic understanding.
- Collaborated with cross-functional engineering teams to integrate AI components into internal systems, gathering user feedback for iterative improvements to LLM applications.
- Maintained clean, testable Python codebase following software engineering best practices for production AI systems.

Tech Stack: Python, LangChain, ChromaDB, Neo4j, SentenceTransformers, RAGAS, Pandas, NumPy, Prompt Engineering, Git.

Education

Brandenburgische Technische Universität Cottbus-Senftenberg **Cottbus, Germany**
Master of Science in Artificial Intelligence *10/2022 – present*

Key subjects:: Machine Learning, Deep Learning, Data Mining, Information Retrieval, Explainable ML, Computer Vision

Mahatma Gandhi University **Kottayam, India**
Bachelor of Computer Applications *2018 – 2021*

Projects

Financial Analysis AI System:

GitHub:: `https://github.com/alwinpaul1/Financial_Crew`

- Architected a **4-agent multi-agent LLM system** with specialized agents for query parsing, code generation, execution, and review, demonstrating end-to-end AI pipeline design.
- Implemented **prompt engineering strategies** and agent orchestration using CrewAI framework to optimize task delegation and response quality.
- Integrated Ollama models with **custom evaluation metrics** to assess code correctness, execution safety, and output quality.
- Developed production-ready **FastMCP server integration** for seamless deployment in AI assistant environments.

SmartTuner: LLM Fine-tuning & Evaluation System:

GitHub:: `https://github.com/alwinpaul1/SmartTuner`

- Developed comprehensive **LLM training and evaluation pipeline** implementing GRPO reinforcement learning for small language models (135M-600M parameters).
- Created **synthetic data generation workflows** using OpenAI API to produce high-quality training datasets for logical reasoning tasks.
- Built **rigorous evaluation framework** measuring model performance across correctness, reasoning quality, and task-specific metrics, achieving 46% $\rightarrow$ 60%+ accuracy improvements.
- Implemented **supervised fine-tuning with LoRA adapters** for parameter-efficient training and deployed real-time monitoring dashboards.
- Wrote clean, maintainable Python code with modular CLI system and comprehensive logging for production deployment.

Job Search Automation System:

GitHub:: `https://github.com/alwinpaul1/Job-Search-TG`

- Engineered **automated data parsing and ingestion pipeline** with Selenium for continuous job data collection and processing.
- Implemented **NLP-based categorization** and deployed notification system with 24/7 operation and error handling.
- Demonstrated proficiency in **REST API integration, workflow automation, and production deployment**.

AI/ML Football Analysis System:

GitHub:: `https://github.com/alwinpaul1/AI-ML-Football-Analysis-System`

- Implemented **computer vision pipeline** with YOLO object detection and ByteTrack for video analysis.
- Applied **data analysis techniques** using Pandas and NumPy for performance metric calculation and optimization.

Achievements

Google Cloud Program: 08/2020 – 11/2020

- Completed advanced training in cloud computing, ML model deployment, and distributed AI systems.

Technical Skills

- **LLM & RAG Development:** RAG Pipeline Design, Prompt Engineering, Chunking Strategies, Data Parsing & Ingestion, LLM Evaluation (RAGAS), Synthetic Data Generation
- **AI Frameworks & Tools:** LangChain, PyTorch, TensorFlow, Transformers, Hugging Face, CrewAI, Ollama, SentenceTransformers
- **Vector & Graph Databases:** ChromaDB, Neo4j, Pinecone, FAISS, Vector Search Optimization
- **Programming & Development:** Python (Expert), Git, Docker, Linux, RESTful APIs, Clean Code Practices
- **Data Science & Analysis:** Pandas, NumPy, Feature Engineering, Data Visualization, Evaluation Metrics
- **Machine Learning:** Reinforcement Learning, Computer Vision, Model Fine-tuning, Deployment Optimization
- **Collaboration:** Cross-functional Teamwork, Agile Workflows, Technical Communication, User Feedback Integration

Languages

German (A2, improving), English (C1)