

Alwin Paul

Cottbus – Germany

☎ +49 176 77354827 • ✉ alwin.paulpv@gmail.com • 🌐 alwinpaul.me
in alwin-paul • 🐙 alwinpaul1

Objective

AI Engineer specializing in **LLM systems, RAG pipelines, and intelligent information retrieval**. Experienced in **LangChain orchestration, prompt optimization, and vector database architecture**. Passionate about building **production-grade AI systems** with measurable performance and real-world impact. Dedicated to deploying scalable language model solutions that drive business value.

Professional Experience

Perinet GmbH

Cottbus, Germany

Working Student – AI & Systems Engineer

07/2024 – present

- Architected end-to-end **Graph RAG pipeline** with **Neo4j knowledge graphs**, combining adaptive PDF chunking with entity-relationship modeling for context-aware document retrieval.
- Built **hybrid retrieval system** integrating **ChromaDB** vector search, **cross-encoder reranking**, and **semantic caching** (0.85 similarity threshold) to reduce latency and retrieval noise.
- Engineered **prompt optimization workflows** with quantitative evaluation metrics, leveraging **RAGAS framework** to benchmark faithfulness, relevancy, and context precision.
- Developed production-grade **LLM evaluation pipelines** measuring hallucination rates, response quality, and retrieval accuracy across multiple RAG configurations.

Tech Stack: Python, LangChain, ChromaDB, Neo4j, SentenceTransformers, Cross-Encoder Rerankers, 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

SmartTuner – GRPO Reinforcement Learning System:

GitHub:: <https://github.com/alwinpaul1/SmartTuner>

- Implemented **Group Relative Policy Optimization (GRPO)** from scratch for training small language models (135M–600M parameters) on logical reasoning tasks.
- Developed **reinforcement learning pipeline** with experience collection, advantage estimation, and PPO-clipped surrogate loss, improving accuracy from **46% to 60%**.
- Built **supervised fine-tuning system** using **LoRA adapters** for parameter-efficient training, generating synthetic reasoning data via OpenAI API.
- Engineered modular **CLI system** with configurable hyperparameters for reproducible experiments and real-time training visualization.

AI/ML Football Analysis System:

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

- Implemented **YOLO object detection** with **ByteTrack** for multi-player tracking across video frames, applying computer vision to sports analytics.
- Utilized **K-Means clustering** for team segmentation and **Optical Flow** for movement analysis, demonstrating versatility in algorithm selection.
- Developed **ball possession tracking** logic determining which player possesses the ball, computing team ball control percentages with interpolation handling for missing detections to ensure temporal consistency.
- Applied **perspective transformation** and **camera motion compensation** to convert pixel coordinates to real-world measurements, enabling accurate speed and distance calculations for player performance evaluation.

Financial Analysis AI System:

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

- Designed **4-agent CrewAI system** (Query Parser, Code Writer, Code Executor, Code Reviewer) for automated financial analysis workflows.
- Integrated **FastMCP server** for **Cursor AI assistant** interoperability, creating extensible tools following industry-standard MCP protocol.
- Developed NLP pipeline using **Ollama's deepseek-r1:7b** to parse stock queries and generate production-ready Python analytics code.
- Leveraged **yfinance API** and **Matplotlib** to build real-time stock analysis and visualization capabilities.

Job Search Automation System:

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

- Built **Telegram bot** with LinkedIn scraping using **BeautifulSoup** and **Requests** for automated job data extraction and parsing.
- Implemented **AI-powered relevance scoring** using **JobBERT sentence transformers** and **TF-IDF classifiers** for semantic job matching.
- Developed **automated alert system** with **APScheduler** for background monitoring (30-min intervals) and instant Telegram notifications.
- Designed **SQLite persistence layer** with connection pooling for user preferences, alert management, and search history tracking.

Achievements

Google Cloud Skills Program – Silver League: 08/2020 – 10/2021

- Earned **43 skill badges** (12,800+ points) with focus on **BigQuery ML**, **Machine Learning APIs**, and **ML data pipelines** on GCP.
- Demonstrated proficiency in **CI/CD workflows**, **Kubernetes deployments**, and **Cloud Observability** (Monitoring/Logging) on GCP.

Technical Skills

- **LLM Development:** LangChain, Prompt Engineering, RAG Design, Chunk Optimization, Context Evaluation, LLM Monitoring
- **Evaluation & Analysis:** RAGAS, Synthetic Data Generation, LLM Benchmarking, Faithfulness & Relevance Scoring
- **AI Frameworks:** PyTorch, TensorFlow, Transformers, SentenceTransformers, Hugging Face, CrewAI, Ollama
- **Vector & Graph Databases:** ChromaDB, Pinecone, Neo4j, FAISS
- **Programming:** Python (Expert), Go, Git, Docker, Linux, RESTful APIs
- **Data Science:** Pandas, NumPy, Scikit-learn, Matplotlib
- **Machine Learning:** Supervised Learning, Reinforcement Learning, Model Fine-tuning, Explainable AI
- **Collaboration:** Cross-functional Development, Agile, Technical Documentation, Research Integration

Languages

English (C1), German (A2, improving)