Alwin Paul

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

Objective

AI Master's student with robust AI knowledge, practical programming skills, and a record of thriving in collaborative environments. Enthusiastic about **machine learning**, **data analysis**, and **system security**, seeking a Master's thesis in **Generative AI** that leverages expertise in Python, ML, and data analysis.

Professional Experience

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

- Explored **GPT model deployment** on resource-constrained ARMv7 Cortex-A8/A9 devices; evaluated frameworks like llama.cpp, TinyLlama, and WebAssembly (WASM) for on-device inference.
- Analyzed architectural constraints and software incompatibilities of 32-bit ARM architectures.
- Developed an **intent-based chatbot** using Natural Language Processing (NLP) techniques as an alternative solution under hardware limitations.
- Created a **Retrieval-Augmented Generation (RAG) system** using Python, Ollama, and Llama model.
- Integrated company product documentation into the RAG system, establishing a knowledge base for efficient information retrieval and enabling query-based access to product information.

Tech Stack: Python, Ollama, Llama, NLP, RAG, GPT, llama.cpp, TinyLlama, WebAssembly, ARM Architecture.

Education

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

Key subjects:: Data Mining, Deep Learning, Data Warehouse, Information Retrieval, Explainable Machine Learning, Image Processing and Computer Vision

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

Key subjects:: Data Structures, Database Management Systems, Design and Analysis of Algorithm, Software Engineering, Computer Networks

Projects

SmartTuner: GRPO Reinforcement Learning System for Small Language Models:

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

- Implemented **Group Relative Policy Optimization (GRPO) algorithm** from scratch for training small language models (135M-600M parameters) to perform logical reasoning tasks.
- Developed complete **RL pipeline** with experience collection, advantage calculation, and PPO clipped surrogate loss implementation achieving **46% → 60%+ accuracy improvements**.
- Built **supervised fine-tuning (SFT) system** with LoRA adapters for parameter-efficient training, generating synthetic reasoning data via OpenAI API integration.
- Created comprehensive **visualization suite** with real-time training monitoring and engineered modular CLI system with configurable hyperparameters for reproducible ML experiments.

AI/ML Football Analysis System:

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

- Implemented **YOLO-based object detection** and ByteTrack for robust tracking, demonstrating ability to apply AI techniques to complex real-world scenarios.
- Utilized **K-Means clustering** and **Optical Flow** techniques, showcasing proficiency in applying various AI algorithms to solve multifaceted problems.
- Developed performance metric modules, indicating capability to derive and implement methods for optimization in dynamic environments.

Explainable AI Quality Inspection System for Industrial Manufacturing:

GitHub:: `https://github.com/alwinpaul1/explainable-ai-quality-inspection`

- Developed custom **CNN architecture** achieving **99.44% test accuracy** for manufacturing defect detection.
- Integrated multiple **XAI techniques** (LIME, SHAP, Grad-CAM) for model interpretability.
- Engineered **data augmentation pipeline** with 9 transformation techniques for balanced training.
- Containerized complete AI pipeline with **Docker** for reproducible cross-platform deployment.

Financial Analysis AI System:

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

- Designed and implemented a **4-agent CrewAI system** (Query Parser, Code Writer, Code Execution, Code Reviewer) for complex financial analysis workflows, demonstrating advanced AI architecture design skills.
- Integrated **FastMCP server** functionality for seamless Cursor AI assistant integration, creating extensible tools that follow industry-standard AI development protocols.
- Developed **NLP pipeline** using Ollama's deepseek-r1:7b model to parse stock queries and generate production Python code for financial analysis.
- Utilized **yfinance API** and matplotlib to create comprehensive stock analysis tools with real-time data processing and interactive visualization capabilities.

Achievements

Google Cloud Program: 08/2020 – 11/2020

- Demonstrated understanding of high-scale applications such as Google Search, Gmail, and YouTube as a result of the program.
- Developed a comprehensive grasp of core concepts, including computing, application development, big data, and machine learning in a cloud environment.

Skills

- **Programming & ML**: Python, PyTorch, TensorFlow, Transformers, Hugging Face, PEFT (LoRA), Scikit-learn, Pandas, NumPy, OpenCV
- **AI/ML Specialization**: Reinforcement Learning (PPO, GRPO), Language Model Fine-tuning, Computer Vision, Generative AI, GANs
- **Data & Visualization**: Matplotlib, Seaborn, Plotly, Training Analytics, Real-time Monitoring, Data Pipeline Development
- **Tools & Systems**: Git, Docker, CLI Development, JSON/YAML Configuration, MySQL, PostgreSQL, Linux, MacOS, Windows
- **Specialized Knowledge**: ML Experiment Design, Research Implementation, Agile Development, Parameter-Efficient Training

Languages Known

English (C1), German (A2)