

AYUSH SINGH

ML Engineer | AI Engineer | Data Scientist

+49 1525 9857962 • ayushrajputsingh.08@gmail.com • www.linkedin.com/in/ayush-singh-1330c • Berlin, DE

Summary

Results-driven Machine Learning Engineer with 4+ years of experience in AI, NLP, and business intelligence, specializing in large language models (LLMs), agentic workflows, and data-driven product development. Proven success in building multi-agent systems and LLM-powered automation pipelines that improve operational efficiency and user engagement. Adept in deploying AI solutions using LangChain, Hugging Face, and cloud-native tools across domains including education tech, market research, and digital marketing, with a strong foundation in Python, MLOps, and end-to-end AI system design.

Project

Project

01/2024 - Present

- AI Research Assistant with RAG & Agents:** Developed a smart assistant using LangChain, LangGraph, OpenAI LLMs, and vector search to answer complex queries with agent-based reasoning and document retrieval.
- Designed custom workflows with **LlamaIndex**, implementing advanced retrieval strategies. Developed a ReAct agent from scratch, solving complex dynamic-k (no. of chunks)
- Built and integrated **AI agents** with **LangGraph** and **LangFlow**, incorporating both short-term and long-term memory capabilities using **Mem0**, enabling adaptive and context-aware interactions
- Synthetic Data Generator with LLMs:** Built a tool using OpenAI's LLMs to generate schema-aligned synthetic data with label conditioning for ML model training.
- Domain-Specific LLM Fine-Tuning:** Fine-tuned an open-source LLM using LoRA on niche domain data, with custom tokenization and evaluation for improved domain-specific performance.
- Created RAG evaluation frameworks across multiple domains, utilizing **RAGAS** and **Langsmith** to assess and optimize Retrieval-Augmented Generation (RAG) performance. Developed custom evaluation datasets for each client
- Deployed a **Llama 3 90b model** on **Google Vertex AI**, focusing on domain-specific synthetic data generation to power various use cases and solution

Experience



Grand View Research

US

Senior ML Engineer

09/2020 - 12/2023

- AUTO ML:** Created End to End Auto **ML orchestration Platform** with UI. enabled data upload from **S3 bucket**, train deep learning models with hyperparameter tuning (**PyTorch**) and monitoring (via **TensorBoard**)
- AI Image Culling software**, implementing a face quality algorithm (MediaPipe) and removing duplicates, blurriness, background removal and blinking images
- Architected a researcher platform and develop custom-built **Federated Machine Learning** and Analytics Python packages.
- Managed **Kubernetes clusters** for containerizing ML models and deploying them to production. Reduced deployment time by 80%.
- Designed an **automated Facebook ad** system with **GPT-3**, boosting engagement rates by 35%
- Developed and implemented customer segmentation using **K-Means clustering** with **Python** and **Scikit-learn**, enabling data-driven targeting strategies and improving marketing effectiveness by **20%**
- Implemented email automation processes with **n8n** reducing errors and costs by 20%, that increased newsletter sign-ups by 15%
- Designed and developed data pipelines for **ETL processes** using **SQL**, Python, and **Apache Spark**.

Education



IU Internationale Hochschule

Berlin, DE

Master of Science in Data Science | GPA: **1.9** / 5

04/2023 - 04/2025

SKILLS

AI/ML Framework: LangChain • LangGraph • LangFlow • Llamaindex • CrewAI • PyTorch • spaCy • Prompt Engineering • RAG

Data Engineering: ETL Pipeline • Data Modeling • Apache Airflow • Fast API • PowerBi • Dash • Plotly • DBT

Cloud, & MLOps: Azure ML • Docker • Kubernetes • Terraform • CI/CD • AWS SageMaker • AWS Lambda • GCP Vertex AI • Azure Data Lake • Azure SQL DB

Database: PostgreSQL • ChromaDB • FAISS • Pinecone • BigQuery

Tools & Libraries: Python • pandas • NumPy • Scikit-learn • Async • pydantic

Languages

English Native

German Intermediate