MUHAMMAD **SULEMAN**

AI Engineer | Data Scientist | Python | LLMs | NLP | Computer Vision | RESTful API | AWS

90763 Fürth, Germany | +49 163 5024790
m.suleman.me96@gmail.com | github.com/Suleman96/
linkedin.com/in/suleman96/ | suleman96.github.io/portfolio/

SUMMARY

AI Engineer specializing in LLMs, AI Automation (N8N-RAG, Lanchain, and more), NLPs computer vision with hands-on expertise in deploying scalable AI applications using cloud platforms (AWS). Experienced in building scalable NLP systems with LLMs and retrieval-augmented generation (RAG). Proven ability to develop real-time intelligent systems, optimize deep learning pipelines, and build end-to-end prototypes for industry-grade deployment. Seeking to contribute to cutting-edge AI product development in a growth-oriented organization.

SKILLS

Programming Languages	**Python**, SQL, C++, MATLAB, Latex, CSS, HTML, Bash
AI / Machine Learning	LLMs (GPT-4, Claude 3.7), Prompt Engineering, **RAG**, Hugging Face, TensorFlow, **RESful APIs**, **Fast API, AI Agent Workflow (Langchain, Cursor),** Fine Tuning**, Performance Evaluation**
Data & Cloud Tools	**AWS**, Docker, VirtualBox, **MLOps**, MySQL, PostgreSQL, Azure ML (Basic familiarity)
Engineering Software	**Tableau**, ROS, Gazebo, MIR, **Power BI**, **Microsoft Office Certified** (Excel, PowerPoint, Word)
Libraries & Frameworks	**PyTorch**, Scikit-learn, OpenCV, Kivy, NumPy, Pandas
Technical Skills	Data Cleaning, Data Analysis, **CI/CD (GitHub Actions)**, Agile, Notion, Data Pipelining
Soft Tools	Communications, Analytical Analysis, Adaptable, Fast Learner, Client Satisfaction

LANGUAGE SKILL

ENGLISH- Fluent **GERMAN**- B1/B2 – Intermediate (willing to improve to C1 for job)

EXPERIENCE

March 2025- Present

AI Product Intelligence Engineer (Working Student)

Astek Heimtextil Gmbh- Fürth, *Germany*

- Created detailed system architecture flowchart, accelerating project alignment and reducing feedback cycles.
- Built a FastAPI ingestion layer to securely receive Shopify product webhooks and batch-import items.
- Designed and implemented a Celery + Redis message queue pipeline for parallel text (GPT-4 API) and image (GPT-4 Vision API) enrichment.
- Built and managed **data ingestion workflows** integrating product feeds via APIs.
- **Collaborated directly with client (CEO)** to gather feedback, resolve issues, and tailor solutions for maximum **client satisfaction**.

Tools: Python, Celery, RESTful API, FastAPI, GPT4, Client Satisfaction

Jan 2025- Present

Artificial Intelligence Engineer Intern

Diplotech Solutions- *Berlin, Germany*

- Developed an intelligent chatbot using **Google Dialogflow ES**, improving client engagement and **reducing first-response time by over 90%**
- Designed a user-friendly chatbot interface with HTML/CSS.
- Deployed via **AWS EC2**, maintaining **99.9% uptime** and ensuring **scalable 24/7 customer support**.
- Proactively addressed client feedback to optimize deployment and ensure smooth user adoption.
- Built a **semantic recommendation app**, leveraging LLM-based text embeddings, **FAISS vector search**, and the RESTful Google Books **API** for **RAG** System.

Tools: AWS EC2, Google API, **RESTful** API, Prompt Engineering, RAG, Google Dialogflow, Gradio

Dec 2023-Sept 2024

Research Student - Master Thesis

Technische Hochschule Deggendorf - *Freyung, Germany*

- **Title:** Dynamic Environment Perception for Autonomous System: Real-Time Approach to Obstacle Detection (**Grade 1.3**)
- Conducted meta-analysis of segmentation and relative depth estimation models.
- Collected and processed datasets for training segmentation models.
- Developed an obstacle detection system for autonomous lawn mower on NVIDIA Jetson Nano.
- Successfully optimized inference speed: reduced latency by 52% (2.25x detection speed)

Tools: Transformers, Segmentation, Depth Estimation, Latex, TensorRT, Linux

May 2023- Nov 2023

Student Research Assistant

Technische Hochschule Deggendorf - *Freyung, Germany*

- **Project HAUSL: Developed real-time obstacle detection algorithms using monocular vision.**
- Performed intrinsic and extrinsic camera calibration and real-time visual mapping.
- Evaluated and implemented machine learning techniques, including ORB, SIFT, and YOLOv8 for robust feature detection.
- Development and augmentation of test datasets encompassing potential obstacle features for robust visual detection.

Tools: Yolov8, cv2, Depth Estimation, Image Detection, Segmentation

Dec 2020 – April 2022

Research Assistant

Digital Pakistan Lab (DPL) - *Islamabad, Pakistan*

- Designed, developed, and documented industrial prototypes (Product Design & Development)
- Conducted structural simulations and 3D prototyping to optimize designs for embedded applications.

Tools: SolidWorks, ANSYS, Cura (3D Printing), Documentation, Product Development

EDUCATION

March 2022 – Sept 2025

M.Eng. Artificial Intelligence for Smart Sensors & Actuator

Technische Hochschule Deggendorf - *Cham, Germany* - Grade: 1.7
Specialization: Machine Learning, Computer Vision, Software Development, Data Engineering

Sept 2015 – July 2019

Bachelor's in Mechanical Engineering
National University of Sciences and Technology (NUST) - *Islamabad, Pakistan*
Specialization: Product Design and Simulation, Structural Analysis, C++ Programming

PROJECTS

- **AI-Powered Book Recommendation System with RAG and LLM Integration**
  - Developed an LLM-powered book recommendation web app using Gradio, FAISS-based (vector indexing) RAG, and Google Books API for intelligent, prompt-based local and external search.
  - **Tools/Techniques**: LLMs, Gradio, FAISS, RAG, Google Books API, RESTful APIs, Prompt Engineering

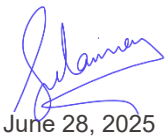
- **AI Driven Adjustment of Public Images/Video for Compliance to General Data Protection Regulation (GDPR)**
  - Built an object detection and tracking model using MobileNet and YOLOv7, annotated through Roboflow, enabling automated privacy blurring for live webcams, images, and videos.
  - **Tools/Techniques:** Roboflow, Mobile-Net, Yolov7, Python

- **Personalized Fitness Recommender App**
  - Developed a Kivy-based mobile app using LSTM to capture, analyze, and visualize user movements.
  - Stored and managed user data effectively using a PostgreSQL database.
  - **Tools/Techniques:** Kivy, LSTM, PostgreSQL

- **Flood Prediction of Passau**
  - Applied time series models (ARIMA, MLP) on tabular time series historical water level and precipitation data for flood forecasting.
  - Validated models using RMSE, delivering actionable insights for early-warning systems.
  - **Tools/Techniques**: Python, ARIMA, MLP, Pandas, NumPy

CERTIFICATIONS

- The Complete Agentic AI Engineering Course – Udemy (2025)
- AI Automation: Build LLM Apps & AI-Agents with n8n & APIs – Udemy (2025)
- LLM Engineering: Master AI, Large Language Models & Agents – Udemy (2025)
- Python for Computer Vision with OpenCV and Deep Learning – Udemy (2024)
- Python Bootcamp 2021- Build 15 working Applications & Games (2021)
- Introduction to JavaScript- Test Automation (2020)
- Microsoft Office Specialist 2013: Word, Excel, PowerPoint – Microsoft (2017)
- Finalist- 'Finding Innovation and Creative Solutions (FICS)' Competition - NUST (2019)
- Director Graphics - Youth Impact (2018)
- General Secretary of Media and Graphics - SAS NUST (2019)



June 28, 2025