Mohammad Ayaz Alam

(+49) 017684051947 | alam.ayaz47@gmail.com | github.com/rebel47 | linkedin.com/in/ayaz-alam

SKILLS

Generative AI Systems:	Agentic AI, MCP, LangGraph, LangChain, RAG, OpenWebUI, Multi-Agent Systems
Natural Language Processing (NLP):	Fine-tuning, Prompt Engineering, Hugging Face, Transformers
Machine Learning & Analytics:	Supervised/Unsupervised Learning, Feature Engineering, Model Evaluation
Programming & Data Science:	Python (Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, PyTorch, Jupyter)
Web Development & APIs:	FastAPI, Flask, Django, RESTful APIs, HTML, CSS, JavaScript, Bootstrap, Tailwind
Mobile & Backend Services:	Flutter, Firebase, Node.js (Cloud Functions)
Databases & Big Data:	SQL, PostgreSQL, Query Optimization, Firebase
Data Visualization & BI Tools:	Power BI, Plotly, Dash
DevOps & Tools:	Docker, Git/GitHub, CI/CD Pipelines, Linux/Unix
Productivity Tools:	MS Excel (Advanced), MS Word, MS PowerPoint, LaTeX

EXPERIENCE

Siemens Energy, *Master's Thesis - Junior Data Scientist (NLP)* | Erlangen, Germany (Hybrid) Jul 2025 - Present

- Architected production-ready multi-agent AI framework using LangChain v1/LangGraph with PostgreSQL memory persistence, RESTful API (FastAPI + Swagger UI), and multi-LLM provider support (Azure OpenAI, Anthropic, Ollama) with hot-swapping capability and real-time streaming responses.
- Implemented MCP server to orchestrate hierarchical multi-agent system with 6+ specialized agents (orchestrator, planner, architect, coder) using role-specific tools and parent-child coordination patterns for complex workflows.
- Established comprehensive testing infrastructure with 22+ pytest tests covering framework initialization, tool integration, conversation memory, and multi-agent collaboration, ensuring production reliability and maintainability.
- Applied framework to Siemens Energy's TMS: Developed agentic solution for operator support automating log trace analysis, incident report generation, automated fix plan recommendations (config diffs, restart/rollback procedures), and email notifications for proactive system maintenance.
- Built NLP components (retrieval, summarization, entity extraction) with extensible tool registry (9+ capabilities: file operations, notifications, database queries) over TMS knowledge bases, enabling intelligent operator copilots for faster incident resolution and reduced downtime.

Institute for Employment Research, *Machine Learning Researcher / Research Assistant* |
Nuremberg, Germany Nov 2024 - Nov 2025

- Applied ML techniques (XGBoost, Random Forest, SVM, PU learning: Elkan & Noto, Two-Step Strategy, Bagging) to forecast labor market trends and analyze job suitability patterns.
- Built scalable data pipelines for preprocessing and feature engineering on 90M+ records, optimizing for efficient data processing and management.
- Developed predictive models to rank job offers and aid caseworkers in pre-selecting opportunities, addressing class imbalances through weight estimation to improve model robustness.
- Conducted comprehensive data analysis, visualization, and EDA to extract key labor market insights and deliver findings through reports and stakeholder presentations.
- Contributed to research paper writing, proofreading, and technical documentation for academic publications.

BASF, *Data Consultant Intern* | Ludwigshafen am Rhein, Germany Apr 2025 - Jul 2025

- Supported internal clients across business units by building and integrating Generative AI tools into daily workflows.
- Developed a Data Validation AI Agent to analyze structured datasets, generate validation reports, compute quality scores, and run custom validation tests.
- Led the end-to-end event app for Inhouse Consulting (submissions, voting, UI, analytics and access controls).
- Contributed to internal project management tools to streamline team operations and task tracking.

Shape AI, *Business Development Intern* | Delhi, India Mar 2021 - Oct 2021

- Performed data analysis using Power BI, extracting actionable insights from sales data to drive business decisions.
- Developed dashboards and visualizations to clearly communicate key metrics across the organization.
- Designed a data-backed course structure for the winter training program, boosting engagement and enrollment.
- Contributed to a marketing project by analyzing target audience behavior, improving course promotion strategies.

Hoda Engineering Works, *Web Developer Intern* | Delhi, India Sept 2020 - Jan 2021

- Built responsive websites using HTML, CSS, Bootstrap, and JavaScript, ensuring performance and cross-browser support.
- Performed testing and debugging on various devices, improving user experience and accessibility.
- Optimized JavaScript and CSS for faster load times, boosting website interactivity and performance.
- Adopted modern web frameworks and best practices, driving continuous improvements in development processes.

PROJECTS

Multi Agent Project Generator - Autonomous AI Project Builder

- Developed an intelligent multi-agent system using LangGraph and Google Gemini AI that automatically generates complete, production-ready projects from natural language prompts.
- Implemented a Planner–Architect–Coder agent hierarchy for structured planning, architecture design, and code generation with Pydantic-based communication.
- Secured project creation and validation through modular tasks and sandboxed file operations.
- Supports multiple tech stacks with configurable recursion and type-safe structured outputs.

URL: github.com/rebel47/Multi-Agent-Project-Generator

OnDevice SLM App – Privacy-Focused On-Device LLM Inference (Flutter)
- Built a on-device LLM application using Flutter and llama_cpp_dart, enabling real-time usage without cloud APIs.
- Integrated 1B–3B GGUF models (Llama, Phi, Gemma), optimized with 4-bit quantization for efficient mobile inference.
- Designed a modern Material Design 3 interface featuring streaming chat, gradient UI, and intelligent selection.
- Implemented model workflows: SAF-based file handling, pre-bundled assets, large-file support, and external detection.
- Optimized inference and memory safety using Android NDK 27, CMake, quantization-aware operations & efficient access.

URL: GitHub: OnDevice LLM App

Fine-Tuned GPT-2 for Poetry Generation with RL
- Fine-tuned GPT-2 (124M) on a Gutenberg poetry corpus, leveraging Reinforcement Learning (RL) to boost stylistic quality and coherence.
- Designed custom reward functions for rhyme, coherence, and creativity.
- Used Transformers for training with adaptive learning rates and policy-gradient methods.
- Published on the Hugging Face Hub; 3600+ downloads with public inference.

URL: ayazfau/GPT2-124M-poetry-RL

RAG Chatbot for Emulating Chat Tone and Personality
- Built a RAG chatbot with Python, Streamlit, and gemini-2.5-flash API for personalized conversation emulation.
- Integrated FAISS and LangChain for efficient similarity search and context retrieval from WhatsApp chat history.
- Designed a persona-based response system ensuring consistent tone across conversations.
- Developed an intuitive UI with real-time chat, multi-session support, and persona configurations.

URL: persona-chat.streamlit.app

Spento - AI-Powered Expense Tracking Web App
- Built a responsive expense tracker with Gemini API for automated receipt analysis and categorization.
- Developed with HTML, Tailwind CSS, and JavaScript on the frontend and Firebase Cloud Functions for the backend.
- Integrated Firebase Authentication (Google Sign-In) and Firestore for secure data storage and user management.
- Deployed on GitHub Pages with real-time dashboards for expense analytics and summaries.

URL: ilhamlabs-ai.github.io/Spento-Web/

CERTIFICATES

Huggingface - AI Agents Fundamentals - Agents fundamentals, Tools and Action, Agents Workflow
iNeuron.ai - Full Stack Data Analytics - Python, SQL, Snowflake, PowerBI, DAX, Excel
iNeuron.ai - Deep Learning Foundation - Neural Network, Backpropagation, CNN
Kaggle - Intro to Machine Learning - Model Validation, Underfitting-Overfitting, Random Forest
freeCodeCamp - Data Analysis with Python - Numpy, Pandas, Data Cleaning

EDUCATION

Friedrich-Alexander-Universität Erlangen-Nürnberg, *MSc. Data Science* | Erlangen, Germany — Oct 2022 - Present
- Courses: Business Analytics, ML Time Series, ML in Finance, Mathematics of Learning, Linked Data

Guru Gobind Singh Indraprastha University, *Bachelors of Computer Applications* | Delhi, India — Aug 2018 - Sept 2021
- Courses: OOPs, Network Security, DBMS, Web Technologies, Data Warehousing and Data Mining

Languages

English	Professional proficiency
German	Intermediate proficiency
Hindi & Urdu	Native proficiency