

Syed Muhammad Taha Imam

Islamabad | simam.bs22seecs@seecs.edu.pk | 03340518269 | [linkedin.com/in/syed-muhammad-taha-imam](https://www.linkedin.com/in/syed-muhammad-taha-imam)
github.com/Syed-MuhammadTaha | syedtahadev.vercel.app

Skills

Transformer Models | NLP | Computer Vision | Deep Learning | Advance TensorFlow | CI/CD pipelines | Prompt Engineering | LLM Fine-Tuning | Multi-Agentic workflows | RAGs | LLM Deployment | Distributed Training Strategies | Open-source LLM Inference | Docker | vLLM | Apache Airflow

Experience

Machine Learning Intern, CogniMind AI – NSTP, NUST Feb 2025 – Present

- Deployed a dockerized Apache Airflow instance on a virtual machine, improving workflow automation efficiency by 40%.
- Designed 5+ scalable DAGs in Airflow to orchestrate data pipelines, reducing manual intervention by 60%.
- Developed CI/CD pipelines with GitHub Actions and Docker, cutting deployment time by 30%
- Improved VLM document extraction accuracy by 20% using prompt engineering techniques
- Optimized inference pipelines by 10% using parallel and batch processing techniques

Machine Learning Intern, RapidsAI – SINES, NUST Sept 2024 – Dec 2024

- Integrated an RAG-based chatbot into a website using Streamlit, enhancing user engagement by 35%.
- Optimized chatbot responses through Chain-of-Thought prompting, reducing incorrect responses by 50%.
- Cut OpenAI API costs 50% with a multi-model query routing system.
- Designed FastAPI endpoint for chatbot responses, enabling user session management and contextual interactions.
- Built a Twitter scraper that processed 500+ tweets daily, enabling large-scale sentiment analysis.

Projects

Attention Is All You Need paper implementation github.com/transformer

- Trained on the Shakespeare data set containing 1 million characters.
- Utilized character tokenization to generate positional and token embeddings for sequence representation.
- Integrated multi-head attention with masking, batch norm, residual connections, and feed forward layers.
- **Tools Used:** Tensorflow, Numpy

CI/CD pipeline for HuggingFace spaces github.com/mllops

- Developed a basic text summarizer using transformers and streamlit
- Deployed the app on HuggingFace spaces
- Engineered a CI/CD pipeline using Github Actions to build and push file to HuggingFace spaces on updates to GitHub repository
- **Tools Used:** Github Actions, HuggingFace, Streamlit

Authored "Unlocking LLM Potential using Effective Prompt Engineering," medium.com/@tahaML
exploring advanced prompt design strategies to optimize LLM performance.

Participated in the VisionRD Hackathon, developing an action classification model using Vision Transformers for real-time video analysis.

Education

National University of Science and Technology, BS in Computer Science Sept 2022 – Jun 2026

- **Coursework:** Linear Algebra, Probability and Statistics, Machine Learning
- **GPA:** 3.43/4.00