

Abdul Samad Gilal

🌐 abdulsgilal.com | ✉ abdulsgilal@gmail.com | [in /abdulsgilal](https://www.linkedin.com/in/abdulsgilal) | [github /abdulsgilal579](https://github.com/abdulsgilal579)

Profile Introduction

AI/ML Engineer with a Bachelor's in Computer Systems Engineering, focused on building and deploying production-grade AI applications on AWS. Currently engaged in research spanning AI safety, computer vision, AI memory systems, and hallucination mitigation in large language models.

Education

Bachelor of Computer Systems Engineering

Jan 2021 – Dec 2024

Mehran University of Engineering & Technology, Jamshoro, Pakistan

CGPA: 2.95 | Percentage: 70%

Work Experience

AI/ML Engineer – Data Pulse Tech LLC, NASTP Karachi

Oct 2025 – Feb 2026

- Developed a computer vision pipeline using MediaPipe for hand tracking of baseball players, detecting pitch release points and throw mechanics
- Built and evaluated a RAG application using RAGAS to benchmark retrieval accuracy and response quality
- Deployed AI systems on Google Cloud Platform, managing infrastructure for production workloads

Marketing Coordinator – Islamic Education and Research Academy, UK

Jul 2024 – Present

- Developed an internal automation tool for caption generation, reducing manual video editing effort
- Created motion graphics and video enhancement workflows for organizational media content
- Maintained data tracking systems using Google Sheets for operational and content analytics

Research Activities

A Comparative Study of Chain-of-Thought Faithfulness in LLaMA-3.1 8B and LLaMA-4 Scout 17B Using Step Intervention Across Mathematics, Factual, and Commonsense Domains 2026

Manuscript in preparation – Investigates CoT faithfulness in two LLaMA models via step-intervention methodology across three reasoning domains.

Projects

RAG Pipeline with Semantic Caching 🔄

- A high-performance RAG pipeline that extracts text from PDFs, stores vector embeddings in Qdrant, and uses Redis as a semantic cache to reduce redundant LLM calls lowering latency and API cost while using Groq for response generation.
- Redis | Qdrant | Hugging Face Transformers | PyMuPDF | Docker | Python | Groq LLM

VisionAsk — Real-Time Scene & System 🔄

- Built a real-time scene Q&A system using object detection, multi-object tracking, and an LLM backend — enabling natural language queries over live webcam feeds with grounded, accurate responses.
- Python | YOLOv8 | OpenCV | SORT (Kalman Filter) | Groq | Whisper

Magic AI Captions 🔄

- Built an AI-powered video captioning pipeline using OpenAI Whisper for speech recognition, burning captions directly into video via FFmpeg and uploading processed output to AWS S3 with downloadable links.
- OpenAI Whisper | FFmpeg | AWS S3 | FastAPI | Python | Docker

Tech Stack

Python, JavaScript, Go · FastAPI, Flask, Node.js · React, Tailwind CSS

AWS, Docker · LLM APIs, RAG, Vector Databases, Computer Vision, Object Tracking

Certifications

AWS Cloud • Docker • Generative AI Langchain • Machine Learning

Pandas • Version Control Git • Javascript