

# Husain Qasim Ali Saif

UAE Golden Visa Holder — Research ML Engineer specializing in LLM Systems & RAG

📍 Dubai, United Arab Emirates    ✉ saifhus365@gmail.com    ☎ 056 985 9353    in Husain Saif    🌐 saifhus365

## Technical Skills

**LLM Systems:** RAG (LangChain, LlamaIndex), Multi-Agent AI (LangGraph, CrewAI), Prompt Engineering, Hallucination Reduction, Embedding Models (BAAI, ColBERT), Vector DBs (Pinecone, Chroma).

**ML & Research:** Fine-tuning LLMs (LoRA, QLoRA), Computer Vision (YOLOv10, ViT, DETR), Experiment Tracking (W&B, MLflow).

**Engineering & MLOps:** Python, PyTorch, FastAPI, Docker, SQL, React, Git, CI/CD (GitHub Actions), Cloud (Azure, GCP), Redis, Cloud Run.

## Experience

### AI Research Assistant

Max Planck Institute of Geoanthropology

Jena, Germany

May 2024 – Jan 2026

- Shipped production **RAG system** serving 100K+ climate research papers with **30% faster query response** (<2s latency) using BAAI embeddings.
- Deployed **multi-agent LLM system** (LangGraph + Llama-3.1-Nemotron-70B) achieving **37% improvement** in retrieval accuracy through systematic evaluation metrics and ablation testing.
- Built React + FastAPI video retrieval interface with ColBERT embeddings for timestamped semantic search; reduced hallucinations **42%** via grounded retrieval validation and custom evaluation harness.

### Machine Learning Intern

Atrillu Media

Bangalore, India

Jun 2021 – Sep 2021

- Built a personalized content recommendation model, increasing influencer engagement by **30%**.
- Analyzed influencer posting patterns using advanced analytics to optimize scheduling.
- Automated influencer-specific app pipelines, reducing manual overhead.

## Projects

### Replication and Transformation of Medical Imaging Data (RetMed)

RetMed [🔗](#)

- Applied SOTA **Object Detection models** (ViT, YOLOv10, DETR) (as of 2024) to the [VindrCXR](#) [🔗](#) dataset for lung abnormality detection, Beating the top competition scores on Kaggle.
- Tools: Python, PyTorch, MMDet, MMYOLO

### Storytelling AI (Illumulus)

Illumulus [🔗](#)

- Enhanced storytelling AI using RAG on Wikipedia documents, producing more realistic narratives.
- Integrated outputs into **Diffusion Models** ([Stable Diffusion 3](#) [🔗](#)) for visual augmentation.
- Tools: Python, LangChain, Ollama, PyTorch, FastAPI, React

## Education

### Bauhaus Universität Weimar, Germany

MS in Computer Science for Digital Media

Oct 2023 – Jan 2026

- Grade: 1.5 (Very Good, Top 10%)
- Coursework: Machine Learning, NLP, Deep Learning for CV, Formal Methods, Image Analysis, Photogrammetry
- Thesis (IJCAI 2026, primary author):** Multi-agent debate framework for scientific idea generation, achieving **27% gain** over SoTA. [\[PDF\]](#) [🔗](#)

### Veermata Jijabai Technological Institute, Mumbai, India

B.Tech in Computer Engineering

Sep 2018 – Mar 2022

- CGPA: 7.23/10
- Coursework: Data Structures, Algorithms, Linear Algebra, Discrete Mathematics, Machine Learning, Cryptography