

Ahmad Mustafa Anis

+92-314-5617800 | <https://ahmadmustafaanis.github.io/> | [linkedin.com/in/ahmad-anis](https://www.linkedin.com/in/ahmad-anis) | github.com/ahmadmustafaanis | ahmadanis5050@gmail.com

EDUCATION

International Islamic University

Bachelor of Science in Computer Science

Islamabad, Pakistan

September 2018 – September 2022

- CGPA: **3.43/4.0**
- Key Courses: Advanced Computer Programming, Object Oriented Paradigm, Data Structures and Algorithms, Operating Systems, Database Technologies, Artificial Intelligence

AI For Global Goals — OxML

Oxford ML Summer School, MLx Fundamentals

Oxford, UK

May 2024 – Jun 2024

- Key Topics: Mathematics of ML, Factorisation techniques, Optimisation, Representation learning, Generative AI

EXPERIENCE

AI Research Fellow

Fatima Fellowship

Sep 2024 – Current

- Medical RAG for Vision Language Models, advised by [Dr. Wei Peng](#) from **Stanford University**

Deep Learning Engineer

Roll AI

Mar 2024 – Current

- Enhanced plain bokeh to Depth based bokeh for realistic effects using Depth Estimation and Segmentation.
- Working on Deep Learning based Active Speaker Detection.
- Working on automated thumbnail generation using Diffusion models.

Machine Learning Engineer

Red Buffer

April 2022 – Mar 2024

- Improved OCR latency from 0.6 FPS to 16 FPS on the CPU for streaming devices to blur sensitive information.
- Integrated Real-Time Dashboards using Dash and Plotly for CO2 Emissions into a Machine Learning Pipeline.
- Reduced injury risk by achieving 95% accuracy using Computer Vision techniques for a mining ship.
- Boosted Accuracy to 85% from 30% for Logo Recognition during cricket matches using CLIP by OpenAI.
- Created a Fact Verification and Identification Pipeline from articles, fine-tuning open-source LLMs like LLAMA v2, Dolly-3B, GPT-Neo and using MultiThreading for a 200% speed increase.
- Developed a Generic Chat With PDF with Dynamic LLM, indexing, vector DB, Multi-Hop Search, and Smart Prompts, using LLama Index.
- R&D for Pose Estimation models for Illegal Bowling Angle Detection in Cricket.
- Communicated technical and project updates with clients.

AI Fellow

PI School of AI

December 2023 – February 2024

- PI School of AI selects a batch of the best engineers from all over the world to turn them into AI specialists via a 2-month fellowship worth 12,500€.
- Extracting Key Information from long context transcripts using Open Source Large Language Models (**MistralLite**, **Mixtral8x7B**) that can run on consumer hardware.

Software Engineer (Deep Learning and Computer Vision)

WortelAI

Dec 2021 – March 2022

- Created a real-time Search Engine using Deep Learning with CLIP by OpenAI, Qdrant, and ElasticSearch.
- Performed Multi-label Text Classification using BERT and PyTorch for 1500+ labels.

Deep Learning and Computer Vision Intern

WortelAI

July 2021 – Dec 2021

- Developed Safety Clothes Detection for Factory workers using Yolov5.
- Trained a Celebrity Detection and Classification system using 3DDFA v2, ResNet50, and a 172GB Dataset.

Contributing Editor

KDNuggets

Sep 2020 – July 2021

- Wrote 30+ original ML articles, accumulating over 410,000 views.
- [Profile Link](#)

PUBLICATIONS

Consent in Crisis: The Rapid Decline of the AI Data Commons

[NeurIPS'24]

- Authors: Shayne Longpre, Robert Mahari, Nikhil Singh, Manuel Cherep, [10 authors], **Ahmad Mustafa Anis**, [30 authors] Daphne Ippolito, Sara Hooker, Jad Kabbara, Alex Pentland

TECHNICAL SKILLS

Programming Languages: Python, C++, Rust, Java, JavaScript

Libraries/Frameworks: PyTorch, Tensorflow, PyTorch Lightning, Distributed Training, OpenCV, Flask, FastAPI

Databases: MySQL, PostgreSQL

Tools: Git, GitHub, GitLab, Docker, Markdown, Jupyter Notebook, Anaconda, Visual Studio Code, Pep8

SELECTED PROJECTS

Musculoskeletal Radiographs Classification

- Detecting normal/abnormal X-rays using the MURA dataset by Stanford, used DenseNet and ResNet-101
- [Code](#)

Multimodal Semantic Search Engine

- A mini search engine that can perform: Text-to-Image, Image-to-Image search, Image-to-text search.
- [Code](#) , [Write-up](#)

DistAYA

- [AYA Expedition](#) project, creating small multilingual LLMs distilled from large multilingual LLMs.
- [Hugging Face](#)

VOLUNTEER EXPERIENCE

Research Contributor | [Data Providence Initiative](#) | [MIT Media Lab](#)

Dec 2023 – Present

- Auditing LLM training data to assess web data access shifts. Measuring cultural representation and advocating for AI transparency.

Reviewer | [MULTILINGUAL REPRESENTATION LEARNING Workshop at EMNLP](#)

Sep 2024 – Oct 2024

- Serving as a Reviewer for 4th Multilingual Representation Learning Workshop at EMNLP 2024

Volunteer Researcher & Regional Community Lead | [Cohere for AI](#)

Nov 2022 – Present

- [Organized over 50 guest speaker sessions with Asian researchers from top universities and institutes.](#)
- Initiated NYU Deep Learning study group (Prof. Yann LeCun).
- Urdu Language Lead for AYA, a Multi-Lingual model by Cohere. Contributed 3 Urdu datasets for AYA collection.

Fellowship Lead | [Bytewise Limited](#)

Mar 2023 – May 2023, Jun 2024 – Aug 2024

- **NLP Lead** (Jun-Aug 2024): Mentoring 20 students/pros, overseeing NLP curriculum.
- **Data Science Lead** (Mar-May 2023): Mentored 50+ students, devised curriculum.

AWS Community Builder | [Amazon Web Services](#)

Aug 2022 – Apr 2023

- Organized events to raise awareness about AWS services in local community.

BLOGS AND TECHNICAL WRITING

[Medium](#), [KDNuggets](#), [CNVRG](#)

INTERVIEWS & PODCASTS

[Interview with AtomCamp](#) discussing the motivation and advice in Machine Learning for beginners

[Podcast with Adventures in Machine Learning](#) how to optimize and improve **OCR** systems

[Podcast with Adventures in Machine Learning](#) on how to deploy Deep Learning models with **FastAPI**

CERTIFICATIONS, ACHIEVEMENTS, & COMMUNITY WORK

Courses & Certifications:

- Microsoft Technology Associate (MTA) 98-381 Introduction to Programming Using Python
- Machine Learning by Coursera
- Deep Learning Specialization by Coursera
- AI for Everyone
- Getting Started with TensorFlow 2 by Coursera
- NYU Deep Learning by Yann LeCun
- Mathematics for Machine Learning by Imperial College London

Achievements:

- [5 kyu on Codewars](#)
- [Platinum badge on KDNuggets](#)
- [2300+ points on StackOverflow](#)