# Arif Khan

Deep Learning | MLOps | GenAI | Vision | AWS | Python | SQL
arif.khan.jmi21@gmail.com | +91-8585970532
LinkedIn | Github

Data Scientist with 5+ years of experience in deploying machine learning models, developing data pipelines, and creating AI-driven solutions. Proficient in Python, AWS, deep learning frameworks, data visualization tools, improving operational efficiency and delivering actionable insights in fast-paced environments.

## EDUCATION

**JAMIA MILLIA ISLAMIA UNIVERSITY** | B.Tech. in Computer Science
July 2020 | New Delhi, India | GPA: 8.3

**MARGARET LEASK MEMORIAL SCHOOL** | Intermediate
May 2016 | Jhansi, India | Percentage : 87%

## EXPERIENCE

**WIN HOME INSPECTION** | Data Scientist 3
Jun 2022 –Present | New Delhi, India

- Successfully created and deployed Flask back-end Remark Search feature for mobile app using AWS Opensearch as a Vector Database, written optimized Elastic-search queries to run on 100k+ documents across 50+ indexes while keeping response time 300ms-900ms.
- Created a deep learning-based data pipeline connecting AWS Simple Queue Service(SQS), AWS EC2 and AWS Opensearch to add/update live data being submitted by 300+ users. Used Pytorch and Transformers model.
- Created our in-house RAG Chatbot, using power of LLM on company internal franchise FAQs and documents.
- Successfully created WIN-AI LLM service using Langchain for 300+ users powered by OpenAI APIs helping users write home inspection remarks in less than 10% of time.
- Implemented a pipeline connecting HubSpot Workflows with AWS Lambda and S3 to generate a Real Estate Agents Database, reducing the cold start of the agents' repository by over 1 year.
- Created a NLP-based repair cost estimation algorithm using transformers.
- Developed multiple internal team dashboards for Inspector Revenue, Ancillary Services, WIN Awards, Google ADs, SEO Data Dashboard using Tableau and Opensearch Kibana Dashboard.
- Migrated highly unstructured data from Umbraco CMS to headless Strapi CMS.
- Implemented an Article Categorization and Content tagging algorithm using NLP to generate tags for each article.
- Created an image classification model to identify appliance tags along with OCR pipeline using Amazon Rekognition API.
- Created Geographical Search algorithm using USA OCRM Engine for WIN and deployed using AWS Lambda and GraphQL serving 10k+ daily searches.

**INSPEKTLABS** | Data Scientist - Computer Vision
May 2019 – June 2022 | New Delhi, India

- Automated Motor Insurance Videos Approval process using Ensemble learning combining multiple object detection and segmentation models reducing TAT from 30 to less than 1 minute.
- Created computer vision models for Cars for detecting damages like dents/scratches/shatters from images and videos.
- Created Segmentation Models for more than 30 unique car parts.
- Made an algorithm to flag out potential claim fraud using AI Models and Ensemble Learning.
- Developed a ReactJS application utilizing AI models to make real-time decisions based on mobile phone camera and magnetometer sensor data.
  May 2019 – June 2020 (Internship)
- Created algorithm to identify key frames from video for faster processing of AI models.
- Created OCR using Google Vision API and AWS Rekognition API to extract key information from Vehicle documents.

**- Alphabots | Data / Back-end Consultant**
Jan 2024 – Present
Helping Alphabots Dev team creating a seamless cloud-based algorithmic trading engine.

**- STTP AI / ML 2024 | Guest Lecturer**
July 2024 | Jamia Millia Islamia University
Conducted Masterclass on Data Analytics and Data Science for 150+ students.

Skills and Tech : Python, AWS, Deep Learning, Tensorflow, PyTorch, Tableau, MYSQL, Flask, REST APIs, GraphQL, OpenAI APIs, Prompt Engineering, Elastic Search, Serverless, Github Actions, CICD, AWS ECS, AWS EC2, AWS Code Pipeline, AWS Redshift, AWS Sagemaker, Google ADs APIs, Google Local Services ADs, PowerBI, Data Scraping.