

# Atharva Niranjana Joshi

[atharvanj@gmail.com](mailto:atharvanj@gmail.com) | London | [linkedin.com/in/atharvanj](https://www.linkedin.com/in/atharvanj) | [atharvanjosshi.github.io/](https://atharvanjosshi.github.io/)

## SUMMARY

---

Senior Data Scientist at KPMG UK with strong expertise in NLP, Generative AI, and LLM-based solutions for misconduct detection. Skilled in Python, Azure, and MLOps, with experience in building production-ready AI pipelines. Holds a Master's in Data Science from Queen Mary University of London and a proven track record of delivering impactful AI-driven insights. Commenced my data career as an AI Consultant at Centre for Health Research and Education using Computer Vision to improve Oral Cancer diagnosis.

## WORK EXPERIENCE

---

### Forensic Data Analyst, KPMG, UK

Oct 2022 – Present

- Actively involved in developing Natural Language Processing (NLP) BERT model using NLTK, Spacy, PyTorch and TensorFlow libraries and TF – IDF Machine Learning models using SGD classifier for unstructured data like text and audio communications to detect Financial and Non-Financial Misconduct in communications reducing 70% of the False Positive outcomes thus saving on manual review time. Used MLFlow for workflow control of the various Machine Learning models.
- Experienced in using Generative AI for synthetic data generation to create Proof of Concept models. Currently working with using Large Language Models (LLM) prompt engineering, Retrieval Augmentation Generation (RAG), LangChain and fine-tuning techniques for models like Llama and GPT to detect specialized risks. Worked with Chroma DB vector database for storage and retrieval of embeddings.
- Experienced in using Microsoft Fabric and its product suite of Data Lakehouse, Warehouse, notebooks, etc to create a fully functional and automated data pipeline.
- Experienced in working in an Agile Team of developers focused on requirements scoping and deployment schedules.
- Used Python extensively for data preprocessing such as audio conversion from raw to usable format, speech/voice to text transcription, extracting useful data from noise, deduplication, language translation and logging and using Git for version control. Used SQL and Microsoft Excel for statistical analysis of structured and unstructured data as well as creating appropriate data visualisation dashboards, model evaluation reports and monitoring data.
- Experienced in MLOps and involved in creating, deploying, operationalizing and maintaining an end-to-end automated CI/CD ETL data pipeline with all stages of preprocessing, model deployment and prediction, and ingestion into Relativity platform thus improving efficiency and saving 10 hours per week.
- Experienced in using Azure Cloud Platform to train and host machine learning and language transcription models on docker containers, using Rest APIs for various services, Azure storage explorer, Azure AI/ML Studio, etc.
- Mentored junior team members and new joiners on the various technical and domain aspects of the project.
- Recognised with 'Employee of the month' award.

### Senior Data Analyst, Centre for Health Research and Education, UK

Jun 2021 – Aug 2022

- Researched and applied various Image object detection and classification models like YOLO v4 and R -CNN using Python and Deep Learning frameworks like PyTorch on existing as well as newly created data.
- Coordinated with various medical institutions and medical experts for obtaining data, managing and storing on Amazon Web Services (AWS) S3 buckets cloud platform for understanding the domain, oral cancer research connecting that with a standalone version of Labelling for annotation to create appropriate datasets.

### Systems Engineer, Infosys Limited, India

Sep 2018 – Sep 2020

- Developed applications in Android and iOS using Xamarin.Forms framework based on C# (Object Oriented Programming) and XAML using DevOps on Microsoft Visual Studio.
- Used Azure Active Directory for authentication and Azure APIs for data retrieval and storage.
- Played an important role in the development cycle by implementing complex features, coordinating with my teammates, tracking the progress of each use case development and bug fixes through the Azure DevOps Portal.
- Developed Pilot projects and also presented Proof of Concepts, the working of apps through emulator and walkthrough of codes to the clients through Microsoft Teams.
- Nominated and awarded with a Certificate of Appreciation for my performance in the company.

## EDUCATION

---

**Master of Science (MSc.),** Big Data Science with Industrial Experience

Queen Mary University of London

Percentage: **86.31 (Distinction)**

**(Key Modules:** *Natural Language Processing, Applied Statistics, Big Data Processing, Data Mining, Machine Learning, Deep Learning & Computer Vision, Neural Networks & NLP, Cloud Computing*)

**Post Graduate Diploma,** Data Science

IMS ProSchool, India

Grade: **A**

**(Key Modules:** *Business Analysis, Advanced Machine Learning and Big Data*)

**Bachelor of Engineering (B.E.),** Information Technology

Savitribai Phule Pune University

Percentage: **63.61 (2:1)**

**(Key Modules:** *Machine Learning, Cloud Computing, Design & Analysis of Algorithms, IT & Project Management, Business Intelligence & Advanced Database*)

**Higher Secondary Certificate (Equivalent to A – Level),** Computer Science

Maharashtra State Board, India

Percentage: **71.69**

**Senior Secondary Certificate (Equivalent to GCSE)**

Maharashtra State Board, India

Percentage: **84.55**

## SKILLS

---

**Data Science:** Python (Pandas, NumPy, Scikit Learn, TensorFlow, PyTorch), SQL, Excel, Relativity, Azure

**Languages:** English (Fluent, C1(IELTS)), Marathi (Fluent, Mother Tongue), Hindi (Intermediate)

## CERTIFICATIONS

---

- Infosys Certified Python Programmer and Python Associate
- Natural Language Processing with Classification and Vector Spaces – Coursera
- Generative AI with Large Language Models – Coursera
- ICA Certification in Financial Crime Prevention
- ICA Certification in KYC and CDD