

Shobhit Maheshwari

shobhit.workds@gmail.com | +91-9599236447 | [Portfolio](#) | [Linkedin](#) | [Github](#) | [Medium](#)

Data Scientist with 6+ years of experience building production AI systems across Computer Vision, NLP, and Generative AI applications. Experienced in developing document intelligence and retrieval-augmented generation (RAG) pipelines using PyTorch, FastAPI, and modern LLM frameworks. Proven track record of deploying ML systems that improve operational efficiency and automate workflows in insurance and information retrieval domains.

SKILLS

Python, SQL, Computer Vision, NLP, Transformers, LLMs, Retrieval-Augmented Generation (RAG), Vector Databases, Embeddings, CNNs, Vision Transformers (ViT), PyTorch, TensorFlow, Scikit-learn, PySpark, LangChain, Keras, FastAPI, Docker, TorchServe, Airflow, Jenkins, AWS, Kubernetes, Nginx, Git

WORK EXPERIENCE

Lead Data Scientist, Roadzen | Remote Apr 2025 - Present

- o Built Autospect, an AI-powered motor insurance claims agent that automated 80% of the pre-inspection workflow, reducing claim intake turnaround time from hours to minutes.
- o Developed a policy parsing pipeline using Qwen-Vision, improving extraction accuracy by 37%, with key business fields achieving up to 90% accuracy.
- o Developed a Gemini-based vehicle identity extraction pipeline with 97.8% accuracy.
- o Developed an LLM-based Text-to-SQL system enabling natural language querying of automotive parts data.

Research Assistant, University of Edinburgh | Edinburgh, UK Aug 2024 - Jan 2025

- o Scraped and processed 70GB of historical FTSE All-Share annual reports for financial QA research.
- o Benchmarked GPT against 4-bit quantized Llama and Mistral models on QA tasks, with GPT outperforming both by ~12% in answer accuracy.
- o Improved QA performance by 16% over baseline RAG pipelines by implementing a MapReduce-based ranking architecture with comparable inference latency.

Senior Data Scientist, Roadzen | Remote Jan 2022 - Aug 2023

- o Built end-to-end ML pipelines for data ingestion, training, and model serving using FastAPI, Airflow, and Jenkins, enabling automated retraining and deployment of in-house ML models.
- o Enhanced an insurance policy QA system using GPT-based retrieval, achieving 95% answer accuracy.
- o Achieved over 84% accuracy in vehicle colour, make, model classification for automated claims validation.

Data Scientist, Roadzen | Remote Jun 2020 - Dec 2021

- o Built a Mask R-CNN model for instance segmentation of vehicle damage, parts, and profiles using PyTorch, reducing claims processing time from 40 minutes to under 2 minutes.
- o Redesigned the mAP evaluation metric to better account for annotation subjectivity, improving model performance achieving 74 mAP.
- o Reduced Damage Recognition API inference latency by 30% using TorchServe and FastAPI.

Data Scientist, Spoonshot | Bangalore, India Jan 2019 - Jun 2020

- o Designed a weighted DeepWalk-based graph model using ingredient co-occurrence and flavour pairing theory to generate novel flavour combinations.
- o Scaled graph edge weight generation to over 100 million ingredient combinations using PySpark.
- o Implemented a Fast R-CNN pipeline for nutrition panel extraction from product images, achieving 85 mAP.

EDUCATION

University of Edinburgh Sept 2023 - Nov 2024
M. Sc. in Data Science (Distinction)

University of Delhi Aug 2015 - Jul 2019
B. Tech. in Information Technology & Mathematical Innovation (89.5/100)