

ALLEN K JOHNSON

alenkj3101@gmail.com | +91 9207685517 | [AlenKJohnson LinkedIn](#) | Kochi, Kerala

SUMMARY

Data Scientist intern and Engineering graduate with experience in developing and deploying AI/ML solutions using both traditional machine learning methods and large language models. Skilled in LangChain-based pipelines, RAG, and fine-tuning to optimize LLMs for domain-specific tasks. Proficient in Python and SQL, with expertise in NLP, deep learning, and model deployment. Passionate about applying advanced AI techniques to deliver scalable and impactful solutions in healthcare and beyond.

WORK EXPERIENCE

DATA SCIENTIST INTERN

April 2025 – Present

PropMarker, Design Direct Web Solutions Ltd

Kochi, India

- Developed image processing, object detection, segmentation, and OCR pipelines using OpenCV for real-world visual data analysis.
- Developed time series models (ARIMA), Python APIs, and worked with MongoDB for data handling.
- Worked on LLM-based projects using RAG, VectorDBs, agents, memory, and multi-agent tools.

DATA SCIENCE AND MACHINE LEARNING INTERN

Aug 2024 – Dec 2024

GeeksForGeeks

Bengaluru, India

- Performed data preprocessing, cleaning, and transformation with Python, Machine learning and SQL for maintaining data integrity.
- Developed and deployed interactive dashboards via Power BI, gaining greater visibility into key performance indicators (KPIs).
- Applied predictive models to improve customer segmentation and streamlined operations, minimizing inefficiencies.

SKILLS

Technical Skills : Machine Learning, Deep Learning, Neural Networks, Predictive Modeling, Natural Language Processing (NLP), Gen AI, Statistical Analysis, Regex, LLMs, RAG, Prompt Engineering

Programming & ML : Python, PyTorch, TensorFlow, Scikit-Learn, Keras, OpenCV, YOLO, HuggingFace, LangChain, Transformers, FAISS, LLAMA, VectorDBs, SQL, MySQL, MongoDB, FastAPI, REST APIs, Git, Docker, Flask, GCP, RAG

Data & Visualization : Power BI, Tableau, Streamlit, Matplotlib, Seaborn, Excel, Pandas, NumPy

Other Tools : Jupyter, VSCode, BeautifulSoup, GNU Octave, MS Office, OpenAI APIs

Soft Skills : Communication, Analytical Thinking, Data Visualization, Team Collaboration, creative problem solving

EDUCATION

Bachelor of Technology in Electronics and Communication

2020 – 2024

APJ Abdul Kalam Technological University

Professional Certificate in Data Science and Machine Learning

Aug 2024–Dec 2024

GeeksForGeeks

Bengaluru, India

PROJECTS

- Vision-Based Floor Plan Analytics and Segmentation Engine (PropMarker, UK):**
Developed 5 OCR and image-processing pipelines, window orientation, dimensions finder, floor separation, passageway detection, outer labels analyser using OpenCV and PaddleOCR, eliminating the need for heavy detection models and streamlining floorplan analysis.
- LLM-powered End-to-End Resume Matcher and ATS Resume Builder:**
AI-powered resume intelligence tool that parses resumes and job descriptions, analyzes skill alignment using LLMs and vector similarity, and tailors ATS-friendly resumes for each role built with FastAPI, Gemini, FAISS, and ReportLab for smart parsing and contextual rewriting.
- Real Estate Price Factor Mining & Prediction Engine (UK Outcodes):**
Developed a web-scraping and NLP-driven system to analyze live real estate factors (amenities, macroeconomic, and geographic) using Serper.dev, Gemini, and MongoDB. The pipeline extracts UK postcode-level insights, summarizes their impact on housing prices, and stores structured results for predictive modeling of property price fluctuations.
- AI-Powered Document Chat and Summarizer:**
Designed a Streamlit application integrating HuggingFace for summarization, FAISS for efficient document indexing, and LLAMA 3.1 for conversational AI. Enables users to upload documents, extract concise summaries, and engage in interactive Q and A with document content.
- AI Powered Medical Query Assistant:**
Developed an AI-powered chatbot leveraging advanced NLP techniques to provide medical guidance. Used SentenceTransformer for semantic embeddings and Groq's API for generating context-aware responses. Deployed a user-friendly interface with Streamlit for seamless interaction.

CERTIFICATIONS & ACHIEVEMENTS

- Tata's GenAI Powered Data Analytics job simulation on Forage
- Data Analysis Using PowerBi, GeeksForGeeks
- Digital Signal Processing, Indian Institute of Technology Palakkad
- VEGA Processor Management Training, Providence College of Engineering