

EXPERIENCE

Turing - AI engineer - backend (Sept’24 - present)

- Project: Working on the backend of LLMs to train them on various datasets. (**end-to-end agent-based data cycle pipeline**)
- methodologies:** My responsibilities include optimizing, and managing prompts for AI and machine learning models. **Collaborate with cross-functional teams to define project requirements and deliver solutions.** Ensure **code quality and maintainability** through code reviews and best practices. **Work on integrating third-party APIs** and libraries into existing systems. Troubleshoot and debug applications to resolve issues and improve performance. Work on **creating RAG and SFT pipelines and incorporating them with RLHF to help fine-tune the LLM.** This significantly improved the **model KPI**.
- Tech stack: LLMs, AWS, databases, Langchain, Python, fast API, ML**

Mindrift - Generative AI engineer (part-time role/freelance) (May’24 - Sept’24)

- Project: **Deploying LLM interfaces using AWS bedrock** ([Detailed demo](#))
- Methodologies:** created an LLM agent-powered question-answering (QA) application over large document corpora using AWS Bedrock and LangChain. I handled challenges like managing large documents exceeding token limits and finding relevant documents. The process involves loading and processing documents, creating vector embeddings with **Amazon Titan**, and storing them in a vector database like SingleStore. When a question is asked to the LLM, relevant document chunks are retrieved and used as context for generating answers with models like **Anthropic Claude through Amazon Bedrock**, following a **Retrieval-Augmented Generation (RAG)** approach.
- Tech Stack: AWS Bedrock, AWS titan, SinglestoreDB, LLMs, RAG**

[National Health Authority \(NHA\)](#) - Backend developer (july’23 - Sep’23) ([Github](#))

- Worked along with a mentor on the **backend framework of the Digital health portal of the National Health Authority**. Part of C4GT (code for gov tech program), a prestigious open-source mentoring program to foster the development of **Digital Public Goods (DPGs)** in India organized by Samagra Governance.
- Implemented the feature of Subscription management flow in ABHA app from scratch**, which helped automate **Subscription Requests**. I emphasized developing and utilizing **microservices** to create a more end-to-end and responsive system. **Using a NoSQL-based database like mongoDB** to store a mapping of different services.
- Tech stack: Springboot, Java, gradle, NoSQL, Docker, Ngrok**

Research Internships

[VLSI Lab IIIT Delhi](#) - Undergrad Research Assistant (Aug’23 - Aug’24)  
Guide: Prof. [Anuj Grover](#)

- Project: Evaluation of the performance of hardware accelerators for implementation of Neural Networks using in-memory computation (**Designing deep learning framework integrated with IMC**)
- methodologies:** Analysing different trained CNNs like Lenet, Alexnet, VGG, GoogLeNet to understand the in-depth layers responsible for most convolution operations and replacing those layers with in-memory computations **to study the improvements in accuracy, execution time, latency, etc. Improved** the execution time **by 3 minutes** on average.
- Tech Stack: Convolutional Neural Networks (CNN), Hardware Architecture, System on a Chip (SoC), Simulink, Keras, TensorFlow**

[MIDAS Lab IIIT Delhi](#) - Undergrad Research Assistant (Dec’21 - Aug’22)  
Guide: Prof. [Rajiv Ratn Shah](#)

- Project: Developed models for image parsing and Data extraction from product images.  
319
- methodologies:** Collected data from over 1000 sample images and organized using labeling studio. Fine-tuned and trained models based on our requirements and Built data quality algorithms based on big data analytics to monitor the data quality. Read papers and implemented various OCR algorithms to find out most accurate and appropriate model for our dataset.
- Tech Stack: Tensorflow, scikit, pandas, seaborn, Python**

EDUCATION

- IIIT Delhi** Jan’2020 - Sept’2024  
New Delhi, India  
*B.tech in Electronics and communications engineering*
- ISG Jubail** 2015-2019  
Jubail, Saudi Arabia  
*American Standards-based curriculum (Accredited by MSA CESS);  
12th percentage: 96.4*

Additional Courses completed:

[Deep learning specialization](#)  
[Tensorflow specialization](#)

PROJECTS

- LLM for resumes** ([Github](#)): This project implements a **Retrieval-Augmented Generation (RAG) pipeline** to process resumes, retrieve relevant information based on user queries, and generate context-aware responses. The system leverages a **combination of vector-based retrieval using embeddings and language models** for response generation. The entire pipeline is integrated into a **REST API for querying** and testing.
- Virtual Try-on chatbot** ([Github](#)): Built a virtual try-on application using **Flask, Twilio's WhatsApp API, and Gradio's virtual try-on model**. Users can send images via WhatsApp to Twilio to try on garments virtually, and the results are returned to them.
- Compliance bot API** ([Github](#)): built an API that takes a webpage as the input and checks the content in the page against a compliance policy then returns the findings (non-compliant results) in the response. **Using Langchain, flask, pydantic, and Open AI API.**
- E-commerce website** ([Github](#)): A headphone e-commerce website created using the MERN stack. Payment gateway made using Stripe API integration. Used Sanity IO for content management and Redux for state management. **Tech stack: HTML/CSS, JS, Nodejs, ReactJS, Vite, NextJs, Sanity IO, Redux, Stripe API, Netlify**

ACHIEVEMENTS AND POSITIONS OF RESPONSIBILITY

- Google Hashcode AIR of under 900 global rank of 2447
- Open source contribution through C4GT (code for govtech) program
- Max rating of 4 stars on CodeChef.
- Leetcode ranking of 1685. 500+ questions solved
- PR Head of ECE department (Jan’2021 - Jan’2022) also managed the department’s website
- Student lead for IEEE IIIT Delhi SB (Feb’21 - Sept’21)
- Student volunteer at NGO called Samaj Sudhar Trust. Worked for 84 hours in total for 6 months.

HOBBIES

- Chess ([Lichess](#))
- MUN
- Blogging (<https://medium.com/@suyash20251>)