# Hasnain Ali

**Full Stack AI Engineer**

**codingwithhasnain@gmail.com || https://www.linkedin.com/in/hasnainali3/ || +923135085477 || Pakistan**

## Profile

**AI/ML Engineer & Full-Stack Developer** with over three years of experience building and launching complete web applications using **React** and **Python (Django)**. Proven ability to architect scalable systems, manage cloud deployments on **AWS**, and integrate cutting-edge **AI/ML** features, including **Generative AI and Large Language Models**, to create intelligent, user-centric products and innovative solutions.

## Skills & Technologies

- **Frontend:** React.js, JavaScript (ES6+), HTML5, CSS3, Tailwind CSS, Bootstrap
- **Backend:** Python (Django, Flask), Node.js (Express), PostgreSQL, MongoDB, RESTful APIs
- **Cloud & DevOps:** AWS (EC2, S3, Lambda), Vercel, Heroku, CI/CD, Docker, Git
- **AI Integration:** LLM Integration (Llama 3, OpenAI), RAG, LangChain, Vector DBs
- **Databases:** PostgreSQL, MongoDB, MySQL, Neo4j

## Experience

### Software Engineer | DEVSINC | OCT 2024 – PRESENT

- Developed full-stack features for healthcare-focused NLP chatbots, building both the Python backend logic and the interactive React-based UI components.
- Built internal tools with React to help evaluate and visualize model outputs from the Turing Apple Project, improving the team's ability to analyze results.
- Integrated RAG architectures to enhance application functionality, bridging backend data retrieval with frontend presentation.

### Associate Software Engineer | AMROOD LABS | MAR 2024 – OCT 2024

- Developed and optimized the primary Django backend for a mobile application, integrating PostgreSQL and AWS S3 for efficient, scalable data storage and retrieval.
- Integrated OpenAI models to automate contract analysis and response generation, cutting manual processing time for legal documents by 60%.

### Founder & Lead Engineer | 2ndPlace *(Jan 2023 - Feb 2024)*

- Led a cross-functional team of 14 to design, build, and launch a full-stack hostel discovery platform from concept to production.
- Architected the complete system using a React frontend and a Django/PostgreSQL backend, supporting thousands of global users and achieving 99.9% uptime via AWS CI/CD.
- Engineered the core search feature with a Google Maps API and a custom REST API, cutting

search latency by 80%.

### FULL-STACK ENGINEER (INTERNSHIP) | ARBISOFT | JUN 2024 – AUG 2024

- Developed a **Django-React** web application with **PostgreSQL**, focusing on full-stack development best practices.
- Optimized **database operations** for faster queries and implemented **bulk data insertion** via load data operation.
- Gained hands-on experience in **frontend development (using React), backend API design, and database optimization**.

## Projects

### 2ndPlace | Hostel Discovery Platform | LINK

**Tech Stack:** Python (Django), TensorFlow, Computer Vision, Google Maps API

- Developed a **hostel discovery platform** that helps users find nearby hostels efficiently.
- Integrated **Google Maps API** for precise location-based searches and **Stripe payments** for seamless transactions.
- Implemented **optimized database solutions** using **PostgreSQL and MongoDB** for scalable data storage.
- Focused on **modern UI/UX design** with best development practices for an intuitive user experience.
- Optimized search functionality, reducing search time by **80%**, enabling users to find hostels 5x faster.

### AutoGen - Client Project

**Tech Stack:** Python, LangChain, PostgreSQL, Vector DBs (ChromaDB), OpenAI API, Gemini API

- Developed a comprehensive framework for building and evaluating Retrieval-Augmented Generation (RAG) systems, focusing on user-controlled performance and security. The system allows non-expert users to deploy custom chatbots based on their proprietary documents.
- **System Architecture & Efficiency:** Designed a scalable, multi-tenant architecture where individual user data and document chunks are isolated. Optimized the retrieval pipeline by allowing users to select specific documents for querying, significantly reducing search space and improving response latency.
- **Model Benchmarking & Configuration:** Implemented a control interface for A/B testing different LLMs (OpenAI, Gemini) and vectorization models. This allows for systematic evaluation of cost, speed, and accuracy trade-offs. Users can configure system parameters, including model temperature and response mode (permissive vs. restrictive), to align with their specific use case.
- **Security & NLP Integration:** Integrated NLP modules for **PII (Personally Identifiable Information) detection** to redact sensitive data before processing and **intent detection** to route queries more efficiently. All conversational data is logged in a structured PostgreSQL database for audit and analysis.

## EDUCATION

**Bachelors in Computer Science**

Comsats University Islamabad (CUI) | CGPA — 3.33