

Shubham Shekhar

Data and Analytics ETL Developer

in.shubhamshekhar@gmail.com

<https://www.linkedin.com/in/shubham-coder/>

<https://shubhamzoro.pythonanywhere.com/>

9122449992

Noida, IN

Data and Analytics Engineer with strong problem-solving skills and a proven ability to learn quickly. Experienced in leveraging AI to automate workflows, optimize data processes, and generate actionable insights. Skilled in building Full Stack projects using Python and React JS. Passionate about solving real-world challenges and committed to continuous improvement, driving mutual growth for both the organization and myself.

Work Experience

Data and Analytics ETL Developer

Present

C3 India Delivery Centre LLP | Noida

- Developed a comprehensive dashboard using Power BI, Power Apps, and Power Automate to evaluate the risk rating of a customer.
- For the same, loaded data from the SQL server into Power BI and then integrated Power BI with Power Apps for write-back capabilities to data sources, and utilized Power Automate to send automated email notifications when changes occur.
- Created SQL scripts to support Actimize CDD implementation, including configuration of standard and customized client-specific risk factors.
- Collaborated with the Center of Excellence team to build a database-integrated chatbot using LangGraph, LangChain, OpenAI, React JS and Python.
- Designed modular agentic workflows for tasks such as memory checks, prompt assistance, schema retrieval, and SQL query execution and built frontend using React JS and backend using Flask.

Data Analyst Intern

Present

C3 India Delivery Centre LLP

- Conducted data processing using Python for making it report-ready.
- Utilised SSMS for writing SQL queries and calculating risk factor scores to assign risk ratings to customers.
- Developed a Power BI report to present intuitive and informative reports effectively .

Projects

SummarizeEase

- Fine-tuned an FLAN-T5 model using PEFT and LoRA techniques on Amazon SageMaker for dialogue summarization tasks.
- Evaluated model performance using ROUGE metrics and optimized outputs through prompt engineering and generative configuration tuning.
- Built an interactive Streamlit web application as a frontend to perform real-time dialogue summarization.

Myntra Clone

- Engineered secure user authentication with JWT tokens and password hashing, protecting sensitive user data.
- Developed RESTful APIs for product catalog, shopping bag, and checkout, fully integrated with Redux-managed frontend state.

- Implemented PostgreSQL full-text search with ranked results and SQL LIKE fallback, enhancing the product discoverability and search performance.
- Deployed frontend on Choreo, backend on Vercel serverless functions, and database on Neon, enabling scalable, low-latency operations.
- Enhanced user experience with async search, responsive design, loading states, and comprehensive error handling.

AI Tutor

- Designed and implemented an AI Tutor application leveraging Python, Flask, LangChain, and Chroma Vector DB for intelligent document retrieval(RAG) and conversational AI.
- Integrated OpenAI GPT-4o-mini for natural language understanding and ElevenLabs STT/TTS APIs to enable real-time speech-to-text and text-to-speech interaction.
- Developed RESTful APIs (/chat, /query, /upload, /stt, /tts) and consumed them with a React frontend, including an animated 3D Mascot using Three.js / React-Three-Fiber.
- Engineered a persistent vector memory system with Chroma for context-aware Q&A across PDF documents(RAG).
- Built voice-enabled conversational flows with speech recognition and synthesis, enabling interactive tutoring experiences.

SkimLit

- Enhanced medical paper readability by categorizing sentences for easier comprehension.
- For the same, I transformed the text data into a structured DataFrame to conduct data cleaning and preparation for model training and classification tasks.
- Employed TensorFlow to construct a classification model using LSTM networks, effectively capturing sentence context and relationships.

Deep Research APP

- Developed a multi-agent AI research application using Streamlit, integrating the OpenAI SDK for advanced language model interactions, and leveraging web search tools to automate comprehensive research, report generation, and email delivery.
- Designed and implemented a scalable AI research platform that utilizes the OpenAI SDK for Intelligent content generation and a robust system of web search tools for data acquisition. Enhancing research efficiency and user experience through a Streamlit interface.
- Engineered an AI-powered research solution featuring automated multi-step queries, a persistent chat history (SQLite), and a dynamic UI, specifically integrating the OpenAI SDK for agentic reasoning and web search tools for Broad information retrieval.

SafeSpace

- Built SafeSpace, an AI-powered mental and physical health assistant using Streamlit, FastAPI, and GPT-4.
- Fine-tuned deepseek-ai/DeepSeek-R1-Distill-Llama-8B on FreedomIntelligence/medical-o1-reasoning- SFT using Unislot and tracked training with Weights & Biases (wandb).
- Used OpenAI embeddings and FAISS for vector-based document retrieval(RAG), enabling contextual medical Q&A from PDF resources.
- Integrated multi-agent routing with LangGraph and delivered empathetic responses using MedGemma, with image diagnosis via GPT-4 Vision.
- Implemented emergency handling by triggering real-time phone calls to predefined contacts using Twilio.

Digital Twin

- Developed a serverless conversational application using AWS Lambda to enable users to interact with a resume via chat, processing user queries dynamically.
- Implemented AWS S3 for hosting a static web interface and storing conversation history, ensuring scalable and durable data storage.
- Configured AWS CloudFront CDN to deliver content globally with low latency and improved performance for the static site.

- Integrated multiple AWS services (Lambda, S3, CloudFront) to create a fully functional, interactive, and secure resume chatbot application.

Core Skills

Python, CPP, Flask, Html, CSS, JavaScript, Bootstrap, MySql, Rest API, Pandas, Scikit-Learn, Numpy, ML, TensorFlow, Generative AI, Amazon SageMaker, Open AI, Hugging Face, Langchain, FastAPI, React JS, ElevenLabs, Power BI, Power Apps, Power Automate, Redux, LangGraph, Vercel, Aws App Runner, Docker, S3, Lambda, CloudWatch

Education

Galgotias College of Engineering and Technology

Bachelor of Technology Computer Science and Engineering
GPA: 7.94

Certificates

100 Days of Code: The Complete Python Pro Bootcamp for 2023

Generative AI with Large Language Models

Machine Learning Specialization

Azure Data Scientist Associate