SHOBHIT KUMAR SINGH

+91-9580727944 shobhit22iit@gmail.com linkedin.com/in/shobhit-singhh github.com/Shobhit-Singhh

EDUCATION

- Indian Institute Of Technology Guwahati, *M.Tech in Medical Devices & Diagnostics* **GPA**: 9.33/10 May 2024
- Kamla Nehru Institute Of Technology, *B.Tech in Mechanical Engineering* | U.P., India **GPA**: 7.75/10 Aug 2021
- Divine Public School, *Mathematics + Computer Science* | Gorakhpur (U.P.), India **GPA**: 7.53/10 July 2017

EXPERIENCE

- **AI Engineer**, *(Neoptio Health Incorporated | Vancouver (BC), Canada)* *(Oct'24 - Current)*
  - **Developed** an advanced psychiatrist chatbot leveraging NLP-LLM techniques to process and interpret complex human responses, enabling the diagnosis of mental health conditions based on **Cognitive Behavioral Therapy (CBT) protocols**.
  - **Engineered** Agent system with **custom RAG, ReAct Prompting , Data Validation, & Instructor models**. Incorporating Advance Prompting techniques like **Meta data gen & Prompt Routing** to enhance reasoning and response precision.
- **AI Tech Lead**, *(FirstBench.Ai | Hyderabad, India)* *(Aug'24 - Jan'25)*
  - Led the AI team for the development of an UPSC Exam Evaluation feature, overseeing the design and implementation from scratch. Features including a Voice Debate feature, a Mock Test feature, and an Essay Evaluation feature using LLM.
  - **Tech Stack:** OpenAi, LangChain, LangGraph, LangSmith, Hugging Face Transformers, spaCy, NLTK, Sk-learn, Pandas.
- **Senior Data Scientist**, *(Primary Healthtech Private Limited | Delhi, India)* *(Feb'23 - Jul'24)*
  - Developed **Python scripts** to streamline data extraction and analysis processes, **resulting in a 50% reduction** in device QA/QC time. We used **hypothesis testing** to **eliminate redundant testing levels 2 & 4**, while **automate procedures for the 7 QC levels**, thus enhancing efficiency and reducing processing time by **90 min / batch**.

DATA SCIENCE PROJECTS & GENERATIVE AI

- **Advance Data Visualisation And Exploration Platform**, *(All-In-One Data Science App)* **Deployed**, **GitHub**
  - Created a **No Code platform** focused on exploratory data analysis and hyper-parameter tuning and model training.
  - **Drag and Drop Tool kit** that simplifies model training and makes machine learning accessible to a broader audience.
  - **Tech Stack:** Python, Numpy, Pandas, Matplotlib, Plotly, Scikit-learn, Pygwalker, Machine learning, Sklearn, Streamlit.
- **Doctor GPT: Healthcare Diagnostic Tool**, *(AI-Based Personal Doctor)* **GitHub**
  - **Fine-tuned T5-base** model using **Name entity recognition** technique to accurately extract symptoms from patient queries.
  - Designed a **RAG agent** that uses **semantic search** to match these extracted symptoms with relevant medical literature.
  - Implemented **multilevel prompting** to analyze risk factors & potential disease cases, provide **personalised patient care**.
  - **Tech Stack:** Hugging Face library, T5, TensorFlow, RAG, Fine-tuned, Multilevel Prompting, NER, wandb, Langchain.

NLP EVOLUTION CHAIN

- **AI-based Question Generation**, *(AI Assistant for EdTech)* **GitHub**
  - Trained a **T5 transformer model** & **quantised** it for text generation, leveraging **BERT to identify contextually similar words** for generating multiple-choice questions, Match-the-Following, True/False questions, & Fill-in-the-Blanks formats.
- **Attention is all you need!** *(Comparing Attention and Self-Attention)* **Github**
  - Trained an **encoder-decoder** model without & with an attention mechanism, & created a transformer model (**positional encoding, multi-head self attention) from scratch** to compare the performance of attention vs self-attention mechanisms.
- **Text Generation Model**, *(Comparing Simple RNN, LSTM, GRU)* **Github**
  - Trained a custom **Word2Vec** model for text embeddings & developed next word prediction models using **Simple RNN, LSTM, & GRU** architectures to evaluate their performance & accuracy with both **high & low dimensional embeddings.**
- **Analyzing Customer Review**, *(Topic Modeling and Sentiment Analysis)* **Github**
  - Utilized Latent Dirichlet Allocation (**LDA**), text processing (**regular expression, stemming & cleaning**), and vectorization methods (**BOW, N-gram, tf-idf**) to analyze sentiment & train a Customer review classification model on a Kaggle dataset.

TECHNICAL SKILLS

- **Programming languages**: *C/C++, Python, MATLAB*.*
- **Development**: *ML, ANN, RNN, NLP, Transformer, LLM.*
- **Libraries**: *Pandas, Numpy, TensorFlow, Sklearn, SpaCy.*
- **MLOps**: *DVC, MLflow, MLops, Docker, Kubernetes.*
- **Deployment**: *Streamlit, Tkinter, FastAPIs, LangChain.*
- **Generative AI**: *LLM, RAG, PEFT, Prompt engineering.*
- **Databases**: *SQL, Vector Database.*
- **Dashboards**: *Tableau, Power BI, TensorBoard, Excel Dashboard.*

COURSES AND CERTIFICATIONS

- **IITG course**: Data structure and algorithm, Database management system, SQL, Mathematical Modelling & Simulation, Image Processing with Machine learning, Supervised and Unsupervised Machine Learning, Python Programming languages.
- **Certification: Stanford online**: Advanced Learning Algorithms, Supervised Machine Learning, **Google**: Foundations of Data Science. Google Advanced Data Analytics Certification, Regression Analysis: Simplify Complex Data Relationships.

POSITION OF RESPONSIBILITY AND EXTRA-CURRICULAR

- **DPR at IITG**: Department Placement Representative for the 2023-2024 batch at the Indian Institute of Technology Guwahati.
- **Coding Milestone**: Solved more than **500 DSA coding questions**, demonstrating proficiency in my problem-solving skills.
- **LeetCode**: Achieved a **rating of 1429 on LeetCode** in weekly contests and have developed a proficiency in problem-solving.
- **Intrapersonal Skills and Interests**: Strength Training, Sketching, Swimming, High-Intensity Interval Training and Gym.
- **Sharing through lens**: Won the **first prize in the 'Shutter UP'** photography contest for my creative visual storytelling.